UNCLASSIFIED

UNCLASSIEIED

# AD-A226 958

DTIC FILE COPY

SEP 26 1990

Approved for public release; distribution Unlimited.

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 90 0943

6a. NAME OF PERFORMING ORGANIZATION
Harvard University

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
Air Force Office of Scientific Research/NL

6c. ADDRESS (City, State and ZIP Code)
Department of Psychology
33 Kirkland Street
Cambridge, MA 02138

7b. ADDRESS (City, State and ZIP Code)
Building 410
Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
AFOSR

8b. OFFICE SYMBOL (If applicable)
NL

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR #89-0461

8c. ADDRESS (City, State and ZIP Code)
Building 410
Bolling AFB, DC 20332-6448

10. SOURCE OF FUNDING NOS.

| PROGRAM ELEMENT NO. | PROJECT NO. | TASK NO. | WORK UNIT NO |
|---|---|---|---|
| 61102f | 2313 | A4 | |

11. TITLE (Include Security Classification)
Perception and the Temporal Properties of Speech

12. PERSONAL AUTHOR(S)
Peter C. Gordon

| 13a. TYPE OF REPORT | 13b. TIME COVERED | 14. DATE OF REPORT (Yr., Mo., Day) | 15. PAGE COUNT |
|---|---|---|---|
| Annual Technical Report | FROM 7/89 TO 7/90 | July 26, 1990 | |

16. SUPPLEMENTARY NOTATION

| COSATI CODES | | |
|---|---|---|
| FIELD | GROUP | SUB. GR. |
| 05 | 09 | |

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
speech perception, prosody, context effects, phonetic segments, fricatives, stress

19. ABSTRACT


**Abstract**

Two series of experiments are reported on the role of prosody in human speech comprehension. One series looked at the role of prosodic information in listeners' ability to recognize adjacent vowels and consonants cued by the common temporal feature of vowel duration. The stimuli consisted of syllables from a large sample of natural speech which listeners heard with prosodic context or without. Prosodic context was found to aid listeners in correctly attributing the phonological source of vowel duration. The second series of experiments examines the role of stress in syllable accessibility during the on-line comprehension of language and from short-term memory. During on-line comprehension stress is found to interact with lexical processing, while the effect of stress on syllable accessibility from short-term memory is not dependent on lexical effects. Partial contents, Disambiguation of (C...)


20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
CLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ OTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION
Unlcassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Dr. Alfred Fregly

22b. TELEPHONE NUMBER (Include Area Code)
(202) 767-5021

22c. OFFICE SYMBOL
NL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

SECURITY CLASSIFICATION OF THIS PAGE

21 AUG 1990

Segmental dependencies by
extended phonetic context
and Coming to terms with
stress / -- / Effects of
stress location in sentence
processing.

## Statement of Work

Work on this project will extend previous work on the context-dependent nature of temporal cues to the identity of phonetic segments, and on the role of coarse-grained aspects of the speech signal in facilitating segment recognition. These extensions will address the following questions: Do adjacent segments exhibit mutual dependencies resulting in perceptual ambiguity that can be overcome by contextual information present in coarse-signal characteristics? Can coarse-grained aspects of the speech signal, lacking sufficient information for segment identification, convey speaking rate independently of variation in the inherent durations of the underlying segments? Do coarse-grained aspects of precursive speech contribute contextual information that is used early in the timecourse of segment recognition? Can coarse-grained aspects of the speech signal direct attention to the location of upcoming stressed syllables?

Work on the project will directly study the nature of coarse-grained aspects of the signal and their relation to processing the suprasegmental temporal aspects of speech. New techniques will be developed for creating coarse-grained representations of speech that eliminate information about segment identity but preserve prosodically-relevant aspects of the speech signal. These techniques will permit control over degree of resolution in the short-time spectrum of speech. Perceptual studies, involving direct judgments on stimuli with varying amounts of spectral resolution, will be performed to determine what the amount of spectral detail that is necessary for perceiving important temporal components of prosody.

As part of the project a computer simulation will be developed that will test the computational adequacy of the processes that are hypothesized to underlie human perception of the temporal properties of speech. This model will address three related issues: the segmentation of speech into syllables, the use of temporal relations between syllables to generate expectancies about the temporal properties of upcoming syllables, and the contextual modulation of feature analyzers for processing temporal cues to segment identity.

## Status of Research

Publications

Eberhardt, J.L., & Gordon, P.C. (1989). The effects of attention on the phonetic integration of acoustic information. **Journal of the Acoustical Society of America**, 86, Suppl. 1.

Gow, D.W., & Gordon, P.C. (1989). Two paradigms for examining the role of phonological stress in sentence processing. **Journal of the Acoustical Society of America**, 86, Suppl. 1.

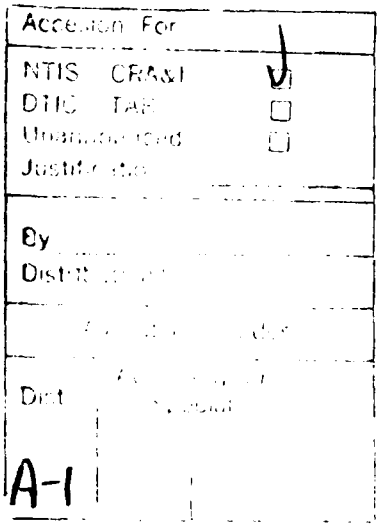
Manuscripts Under Review

Gordon, P.C., Schaeffer, C.P., & Kennison, S.M. Disambiguation of segmental dependencies by extended phonetic context. Manuscript submitted to *Perception & Psychophysics.*

Gow, D.W., & Gordon, P.C. Syllable stress in the processing and representation of spoken sentences. Manuscript submitted to *Journal of Experimental Psychology: Human Perception and Performance.*

Manuscripts in Preparation

Gordon, P.C., Eberhardt, J.L., & Rueckl, J.G. The role of attention in determining the phonetic significance of acoustic cues.




# Disambiguation of Segmental Dependencies by Extended Phonetic Context

Peter C. Gordon  Christopher P. Schaeffer  Shelia M. Kennison

Harvard University

Running Head: Dependencies in Segment Recognition

Send Correspondence to:

Peter C. Gordon
Department of Psychology
Harvard University
33 Kirkland St.
Cambridge, MA 02138
(617) 495-0848

**Abstract**

Two experiments investigated listeners' ability to recognize adjacent vowels and consonants that are conveyed in part by a common temporal cue -- vowel duration. The stimuli consisted of a large sample of natural speech containing nonsense syllables made by combining four vowels that differed in inherent duration (/I/, /ɛ/, /i/, /e/), with two syllable-final consonants that differed in phonological voicing (/z/ and /s/), a distinction that is partially cued by vowel duration. In Experiment 1, listeners identified the syllables after they had been gated from sentential context. Accuracy in recognizing /s/ varied with the inherent duration of the preceding vowel with the number of errors increasing with inherent vowel duration. This suggests that listeners undercompensated for the effect of vowel identity when interpreting vowel duration as a cue to syllable-final voicing. In Experiment 2, listeners heard the syllables in sentential context. Accuracy improved and no relation was found between inherent vowel duration and accuracy in recognizing /s/. This indicates that the prosodic context of a syllable conveys information that is useful to listeners in correctly determining the segmental basis of the internal temporal structure of a syllable.

The temporal structure of speech is determined by many linguistic and non-linguistic factors and is responsible for conveying a multitude of information to the listener. Because of this, the manner in which temporal aspects of the speech signal encode information and in which listeners successfully decode it has been the subject of considerable study (e.g., Fowler, 1980; Gordon, 1988; Klatt, 1976; Miller, 1981; Port, Al-Ani & Maeda, 1980). The present effort extends this study by pursuing two goals: (1) To assess the segmental ambiguity of the temporal characteristics of local stretches of speech by examining the perception of adjacent phonetic segments whose identities are conveyed in part by a common durational cue, and (2) To understand whether such local ambiguity is diminished by perception of the overall prosodic pattern of an utterance. The results of two experiments show that listeners' accuracy in recognizing syllable-final /s/ depends on the inherent durational characteristics of the preceding vowel when syllables are gated from sentential context, but that this dependency is not present when syllables are heard in sentential context. These results are interpreted as indicating that listeners undercompensate for the effects of vowel identity on vowel duration when they can not use the overall temporal context of an utterance. Possible mechanisms are discussed by which perceiving the prosodic context of a syllable might help calibrate a listener's perception of its internal temporal structure, leading to more accurate segment recognition.

The joint effect of adjacent segments on a common durational cue provides a stimulus situation that is potentially quite revealling about how listeners handle ambiguity in the speech signal. Such a situation occurs for the duration of a vowel as it is influenced both by vowel identity and by the phonological voicing of an immediately following consonant. Vowels differ naturally in their inherent durations (Peterson & Lehiste; 1960) and changes in vowel duration provide a sufficient cue to shift a percept from a long vowel (such as /i/ as in "beat") and a short vowel (such as /I/ as in "bit") when the formant pattern is ambiguous (Ainsworth, 1972). The phonological voicing of syllable-final consonants also influences vowel duration, causing the preceding vowel to be shorter when the consonant is voiceless relative to when it is voiced (Denes, 1955; Peterson & Lehiste, 1960);

perceptual experiments verify that vowel duration is a cue to syllable-final voicing (Denes, 1955). Given this common effect of vowel identity and syllable-final consonant voicing, there are three possible patterns that might be observed in the joint effect of vowel duration on perceived vowel length and consonant voicing. First, listeners might *overcompensate* in attributing the value of the common durational cue to one of the segments. This would result in a negative correlation between perceived vowel length and consonant voicing; i.e., long vowels with voiceless consonants and short vowels with voiced consonants. If such a pattern were found, and if one of the segments were recognized more accurately, then we might infer that the more accurately recognized segment is recognized first and that the vowel duration characteristics are attributed to it. Second, listeners might *undercompensate* and attribute the durational effect of one of the segments to the other segment as well. This would produce a positive correlation between vowel length and voicing; i.e., long vowels with voiced consonants and short vowels with voiceless consonants. This pattern might indicate that vowel duration is simultaneously being attributed to both segments. Third, perception of vowel length and consonant voicing might be *independent*; that is the accuracy of recognizing one segment would not depend on the identity of the other. This might happen if cues distinctly related to the individual segments provided enough information to overcome ambiguity in the common durational cue. Independence might also be found if the characterization of vowel duration as an ambiguous cue is wrong or if vowel duration *per se* is not the effective perceptual cue but rather is correlated with some other aspects of the speech signal that unambiguously convey the identity of the segments (cf. Fowler, 1980; Soli, 1982). Determining whether overcompensation, undercompensation, or independence best characterizes speech perception would place an important constraint on models of segment recognition.

The most systematic study of listeners' joint perception of vowel length and consonant voicing was conducted by Mermelstein (1977). His goal was to assess whether phonemes or syllables are the basic unit for perceiving speech. He reasoned that phonemes would be supported as the basic units if perception of vowel length and consonant voicing were found to be independent. On

the other hand, a dependence between vowel length and consonant voicing would suggest the two are perceived interactively as might be expected if whole syllables were the basic units for perceiving speech. In an experiment using synthetic speech varying in frequency of the first formant (F1) and vowel duration, he examined listeners' joint identifications of /æ/ vs. /ɛ/, and /d/ vs. /t/, in the words: /bæd/, /bed/, /bæt/ and /bet/. Mermelstein found that listeners identification of both vowels and consonants depended both on F1 and vowel duration. Thus, there was some correlation overall between the two judgments because they are based on common factors. However, the critical question for Mermelstein was whether the judgment on one segment was related to the judgment on the other segment when stimulus characteristics were held constant. This could only be examined for those stimuli that were not identified with perfect consistency. For those stimuli, Mermelstein found that 6 of 10 subjects showed no dependence between vowel and consonant judgment, two showed a negative correlation and two showed a positive correlation. Mermelstein interpreted these results as supporting a model in which phonemes are basic units of perception that are identified by independent decision processes.

Mermelstein's elegant analysis indicates that vowel and consonant identifications are independent for the situation that he studied. However, the resulting simple model of the relation between successive phonemes is achieved at the cost of a complex view of the relation between acoustic cues and the recognition of an individual phoneme; e.g., the acoustic cues in the vocalic region that support perception of syllable-final consonant voicing must be taken to include F1 frequency of the preceding vowel as well as its duration. In addition, Mermelstein's results are not informative about whether naturally-occuring vowel-consonant sequences unambiguously encode information about vowel and consonant identity in a way that could be recovered by such an independent decision process. A study by Gordon (1989) suggests that they may not. This study examined listeners' accuracy in recognizing the voicing of syllable-final /s/ and /z/ in a relatively large set of naturally produced syllables. In certain conditions, it was found that accuracy in recognizing the fricative depended on the inherent durational characteristics of the preceding vowel. Listeners

appeared to undercompensate for the effects of vowel identity when interpeting the significance of vowel duration as a cue for voicing, leading to a positive correlation between vowel length and perceived fricative voicing. However, because listeners were not asked to identify the vowel, the exact nature of the dependence between vowel and consonant could not be specified (Gordon, 1989). The present experiments more systematically assess whether listeners can recover speakers' phonetic intentions when the identity of successive phonemes is cued in part by a common signal characteristic. It also examines the whether this ability is influenced by the prosodic context of a syllable.

A large number of studies (see Gordon, 1988; 1989; Miller, 1981; 1987; Summerfield, 1981 for discussions) have shown that extended prosodic context influences perception of the local temporal components of speech. Of particular interest to the present study, Gordon (1989) showed that accurate interpretation of vowel duration as a cue to voicing in syllable-final fricatives (/s/ vs. /z/) partly depended on incorporating information about phrase position. Syllables gated from phrase-internal position were more likely to be perceived as ending in /s/, while syllables gated from phrase-final position were more likely to be perceived as /z/. Presumably this is because syllables in phrase-internal position have shorter vowels than those in phrase-final position (Klatt, 1976; Martin, 1970), and listeners needed to take this information into account in interpreting vowel duration as a cue to fricative voicing. This finding shows that listeners adjust their expectations of local speaking rate based on sentential context. The present study further explores the extent to which perceiving the sentential context of a syllable helps listeners calibrate their perception of the internal temporal structure of a syllable.

## Experiment 1

This experiment explores dependencies between perception of vowel identity and syllable-final fricative identity in CVC syllables gated from sentential context. The syllables, which all began

with /t/, could have either long (/e/ or /i/) or short (/ɛ/ or /ɪ/) vowels[1] and a voiced (/z/) or voiceless (/s/) fricative as final consonants. There were three reasons for studying the recognition of these syllables after gating: First, possible dependencies between vowel and consonant can not be analyzed under conditions of perfect recognition, and gating provides a relatively non-disruptive way of inducing errors both for vowels (Verbrugge & Shankweiler, 1977, discussed in Miller, 1981) and for syllable-final fricatives (Gordon, 1989). Second, some of the errors that result from gating appear to represent systematic misinterpretation of temporal components of a syllable (Gordon, 1989; Verbrugge & Shankweiler, 1977, discussed in Miller, 1981) and perception of the temporal components of syllables is the domain of present interest. Third, some of the results of Gordon (1989) suggested that there was a greater effect of inherent vowel length on perception of voicing when syllables were gated as compared to when they were heard in sentential context. These studies show that gating naturally-produced syllables from sentential context can result in systematic misperceptions of segment identity that may be informative about the kinds of temporal information that listeners ordinarily integrate in segment recognition. In the present experiment, subjects were asked to identify both the vowel and the final consonant of gated syllables in order to examine whether systematic dependencies between the two indicate an interaction between the processing of vowels and consonants.

## Method

*Subjects.* Twelve young adults attending classes at Harvard University were paid $5.00 to serve as paid subjects in a single session lasting approximately 50 minutes. They were recruited with posted notices and reported no history of hearing or speech difficulties.

*Stimuli.* The stimuli consisted of eight syllables that varied in their vowels (/e/, /i/, / /, or /ɪ/) and their final consonants (/z/ or /s/). These syllables were spoken in four sentence frames that varied the phrase position of the syllable (internal vs. final), and the phonological voicing of the

consonant immediately succeeding the test syllable. Table 1 shows the eight syllables and four sentence frames.

In combination, the eight syllables and four sentence frames yield 32 sentences. Two repetitions of each of the sentences were spoken by 10 native speakers of American English recruited from the same population as for the listening experiments. The sentences were presented to the speakers on a CRT screen in a random order. Speakers were asked to read each sentence aloud in a natural voice with normal intonation. Prior to engaging in the task, the speakers were drilled with flashcards on the appropriate pronunciations of the orthographic representations of the test syllables. Recordings were made in a sound-attenuating chamber using a Shure SM59 microphone and a Nachamichi BX-100 tape deck.

The resulting 640 sentences were low-pass filtered at 9.7 kHz and digitized at 20kHz. Boundaries for the test syllables were determined and they were gated from their sentence frames. The onset of the syllable was defined as the release of the initial /t/ and the offset was defined as the closure for the following stop.

*Design and procedure.* The 640 test syllables were grouped into 20 blocks of 32 trials. Each block included all combinations of phrase position, following consonant, vowel identity and fricative identity, as well as roughly equal representation of the different speakers. The order of the syllables within each block was randomized and the syllables were output onto audio tape with four seconds of silence between each syllable. Subjects in the listening experiment were told that they would hear a series of syllables and should identify them by circling the appropriate syllable on an answer sheet. The answer sheets listed the eight alternative syllables for each trial. Subjects sat in a sound-attenuating booth and heard the syllables at a comfortable listening level over Sennheisser HD 430 headphones.

## Results

Figure 1 shows listeners' mean accuracy in recognizing the intended fricative as a function of the linguistic manipulations. Given the number of possible main effects and interactions, we will only report significance levels for effects that reach at least a 0.10 level both by listeners ($F_1$) and by speakers ($F_2$). Recognition accuracy for /z/ (96.8%) was substantially higher than for /s/ (78.2 %); $F_1(1,11) = 81.4$, $p < .001$; $F_2(1,9) = 10.6$, $p < .025$. A main effect of vowel identity was also observed; $F_1(3,33) = 21.0$, $p < .001$, $F_2(3,27) = 6.0$, $p < .005$. The effect of vowel identity on fricative recognition can be seen more clearly in light of a significant interaction between vowel identity and fricative identity. Accuracy for /s/ declined with increasing vowel length (88.3%, 80.4%, 74.9% and 69.2%) while accuracy on /z/ was relatively unaffected (96.6%, 97.9%, 96.5% and 96.4%); $F_1(3,33) = 17.1$, $p < .001$, $F_2(3,27) = 5.5$, $p < .005$. A significant interaction of fricative identity and phrase position was observed. Accuracy in recognizing /s/ was better in phrase-internal position (82.8%) than in phrase-final position (73.6%) while the opposite was true for /z/ (98.0% in phrase-final position and 95.7% in phrase-internal position); $F_1(1,11) = 52.5$, $p < .001$; $F_2(1,9) = 9.0$, $p < .025$. The interaction of fricative identity and voicing of the following consonant was significant by listeners, $F_1(1,11) = 67.5$, $p < .001$, and close-to significant by speakers, $F_2(1,9) = 4.8$, $p < .075$. Accuracy tended to be higher for /s/ followed by /t/ (81.0%) than by /d/ (75.4%), while for /z/ accuracy was higher followed by /d/ (97.8%) than by /t/ (95.9%).

Figure 2 shows listeners' mean accuracy in recognizing the intended vowel as a function of the linguistic manipulations. Overall accuracy was very high (96.4%), and no main effects or interactions reached a .10 significance level when tested by speaker. However, a number of effects were found to be significant by listener. These effects will be reported with the caution that they reflect the impact of particular stimuli generated by processes whose generality is not established with the current sample of 10 speakers. A significant interaction was observed between vowel identity and fricative identity; $F_1(3,33) = 3.0$, $p < .05$. However, this effect seems to be unsystematically related to the expected effect of fricative-voicing on vowel duration as it might comprise a cue to vowel identity. A significant interaction of vowel identity and phrase position was observed; $F_1(3,33) = 12.0$, $p < .001$. Here, the interaction does make sense in terms of how phrase

position might influence vowel duration as a cue to vowel identity. The linear interaction of vowel identity (arranged in order of duration, Peterson & Lehiste, 1960) and phrase position was significant; $F_1(1,11) = 9.0$, $p < .025$. Significant interactions were also observed between: phrase position and fricative [$F_1(1,11) = 8.0$, $p < .025$], phrase position, fricative and vowel [$F_1(3,33) = 4.0$, $p < .05$], voicing of following consonant, fricative and vowel [$F_1(3,33) = 4.0$, $p < .05$], and voicing of following consonant, phrase position, fricative and vowel [$F_1(3,33) = 3.0$, $p < .05$].[2]

Discussion

The results of the experiment extend some previous findings concerning the perception of temporal cues to segment identity, while suggesting limits on the generality of some others. With regard to syllable-final fricative recognition, the results support Gordon's (1989) finding that listeners are more accurate when the effects of phrase position on vowel duration are congruent with the effects of fricative voicing. However, the results differ from Gordon (1989) in the finding that recognition accuracy for /z/ is much higher than for /s/. A plausible account of why /z/ would be more accurately recognized than /s/ might appeal to the idea that the voiced-frication interval (the time between the onset of frication and the offset of voicing) could constitute a relatively invariant, positive cue that a segment is voiced (Gordon, 1989) but that the absence of such a cue would not indicate conclusively that a segment was unvoiced. The current experiment differed from Gordon (1989) in that the syllables began with /t/ rather than /w/ or /b/, but it is not obvious why this should matter. The current experiment also used a much larger sample of speakers (10 vs. 3) than Gordon (1989), which allowed results to be generalized over speakers ($F_2$) as well as listeners ($F_1$). The failure of Gordon (1989) to find a main effect of fricative identity may reflect limitations of the stimulus sample in that experiment. However, the result of primary theoretical significance in Gordon (1989) -- the interaction between fricative and phrase position -- was obtained in the present experiment, and was found to be significant by speaker as well as listener. A new finding in the present study was that accuracy in recognizing the voicing of the final consonants of the gated syllables was higher when it was the same as the voicing of the subsequent (and not heard)

consonant. A plausible interpretation of this finding is that there is less devoicing of the fricative when the subsequent consonant is voiced rather than voiceless. A greater voiced-frication interval would promote /z/ percepts, while a shorter one would promote /s/ percepts.

With regard to vowel recognition, the very high accuracies obtained suggest that the syllables contained enough distinctive information that their vowels could be readily identified even when contextual information from the surrounding sentence was not available. In a test by listeners, vowel recognition was significantly higher when the effect of phrase position on vowel duration was congruent with the effect of vowel duration on vowel identity. However, this effect did not generalize across speakers indicating that the present study provides only weak support for the idea that listeners will conflate the temporal effects of phrase position with the temporal effects of vowel identity when identifying vowels gated from context. Verbrugge and Shankweiler (1977; discussed in Miller, 1981) found that listeners mistakenly attributed the effects of overall speaking rate to gated vowels, but that this effect did not extend to the (presumably) smaller durational impact of stress when vowels were gated from context. The absence of a phrase position by vowel length interaction (as tested by speaker) in the present study suggests that vowel recognition may not be dependent on the use of extended context to overcome local speaking rate variation.

The major goal of the experiment was to see whether there were dependencies between vowel recognition and fricative recognition. The results showed no dependency of vowel accuracy on neighboring fricative. It may be the case that no such dependency exists, or that the high recognition accuracy for the vowels obscured a possible dependency. The results did show that fricative accuracy for /s/ was highly dependent on vowel length. Accuracy steadily declined from 88% to 69% as the inherent vowel duration increased, indicating a negative correlation between perceived voicing and vowel length. Such a negative correlation suggests that listeners undercompensated for the effects of inherent vowel duration when identifying the final /s/s. A discussion of possible mechanisms that might produce such an undercompensation will be postponed until after a further assessment is made of the conditions under which it is observed.

## Experiment 2

This experiment examines whether the dependency between fricative voicing and vowel length is observed when the syllables are heard in sentential context. The syllables were gated in the previous experiment in order to induce some misperceptions and because previous research had shown that misperceptions of gated syllables sometimes involved mis-encoding the temporal properties of syllables. A better understanding of the dependency between fricative and vowel perception can be obtained by seeing whether it persists when syllables are heard in the context in which they were produced. If the dependency does persist, it would indicate that successive segments that are conveyed by a common temporal cue are quite difficult to transmitt from speaker to listener. If the dependency is substantially diminished, it would indicate that listeners can use information conveyed by extended phonetic context in order to accurately calibrate their perception of the temporal cues within a syllable.

### Method

*Subjects.* Twelve individuals who had not participated in the previous experiment served as paid subjects in two sessions lasting approximately 50 minutes each.

*Stimuli, design and procedure.* The stimuli were the same as in Experiment 1, except that they were presented in the sentences in which they were spoken. The number of blocks and order of presentation were the same as before, as were the response sheets. There were four seconds of silence between each sentence. Because of the greater duration of each trial, listeners were tested in two sessions instead of one.

### Results

Figure 3 shows listeners' mean accuracy in recognizing the intended fricative as a function of the linguistic manipulations. Only two effects came close to being significant by both listener ($F_1$)

and speaker ($F_2$). Recognition tended to be more accurate when the syllable ended in /z/ (97.6%) than when it ended in /s/ (89.8); $F_1(1,11) = 26.8$, $p < .01$, $F_2(1,9) = 5.1$, $.05 < p < .06$. The interaction of phrase position and fricative was significant by listener, $F_1(1,11) = 9.0$, $p < .05$, and marginally significant by speaker $F_2(1,9) = 4.5$, $p < .10$. In the previous experiment, the strong interaction of fricative identity and vowel identity was the most interesting new result. Examination of Figure 3 shows that in the current experiment the vowel-fricative interaction was much less systematic than before. This is born out by the significance levels as well; $F_1(3,33) = 3.7$, $p < .05$, $F_2(3,27) = 0.7$.

Figure 4 shows listeners' mean accuracy in recognizing the intended vowel as a function of the linguistic manipulations. As in the last experiment, accuracy in recognizing vowel identity was very high (mean accuracy = 97.4%). Only one effect approached significance: The interaction of phrase position by vowel identity was significant by listener, $F_1(3,33) = 10.0$, $p < .01$ and marginally significant by speaker, $F_2(3,27) = 2.5$, $p < .10$. This interaction took the form of greater accuracy for short vowels (/I/ and / /) than long vowels in phrase internal-position, and greater accuracy for long vowels (/i/ and /e/) in phrase-final position.

Discussion

The results of the experiment show that recognition of the segments is relatively uninfluenced by the linguistic relations when the syllables are heard in their sentential context. Those influences that were significant by listener were only marginally significant by speaker. For vowels, accuracy was slightly higher when the effect of phrase position on vowel duration was congruent with vowel length. While the effect was small and only marginally reliable by speaker, this pattern suggests that the presence of sentential context may not be sufficient to perfectly disentangle the two sources that affect vowel duration. For consonants, the accuracy was again higher for /z/ than /s/. Also, phrase position interacted with fricative such that /z/ was more accurately recognized in phrase-final than in phrase-internal position, and /s/ was more accurately recognized in phrase-internal than phrase-final position. Again, this indicates that the presence of

sentential context is not always sufficient to allow listeners to sort out prosodic from segmental sources of variation in duration (a result observed in Gordon, 1989 as well). Beyond these small effects, the most interesting finding of Experiment 2 was that some of the linguistic relations that affected accuracy in Experiment 1 appeared to have little or no influence. In order to substantiate this difference, a set of between-subject analyses were performed comparing the results of the two experiments.

Comparison of Experiments 1 and 2

For vowel accuracy, the between-experiment gating manipulation had no significant main effects or interactions. However, several significant effects were observed for fricative accuracy. Recognition was more accurate when fricatives were heard in context than when they were gated; $F_1(1,22) = 32.2$, $p < .001$, $F_2(1,9) = 19.0$, $p < .005$. Gating interacted with fricative identity, with a greater benefit of context being shown for /s/ than for /z/; $F_1(1,22) = 17.5$, $p < .001$, $F_2(1,9) = 20.0$, $p < .005$. The three-way interaction of gating, fricative identity and phrase position was significant by listener, $F_1(1,22) = 12.2$, $p < .005$, and close-to-significant by speaker, $F_2(1,9) = 4.4$, $p < .075$. The pattern of the interaction was consistent with that observed by Gordon (1989): /s/ benefitted more from the presence of context in phrase-final position while /z/ benefitted more in phrase-internal position. This effect was interpreted by Gordon (1989) as indicating that the presence of context allowed listeners to factor out the effects of phrase position on vowel duration when interpreting it as a cue to the identity of the final fricative. The three-way interaction of gating, fricative identity and following consonant was also significant by listener, $F_1(1,22) = 19.5$, $p < .001$, and close-to significant by speaker, $F_2(1,9) = 4.9$, $p < .075$. The form of this interaction was that /s/ benefitted from context more when the following consonant was voiced (i.e., /d/) while /z/ beneffited most from context when the following consonant was voiceless (i.e., /t/). A likely account of this pattern is that the duration of the voiced-frication interval was influenced by the voicing of the following consonant, so that knowledge of this consonant enabled listeners to accurately factor out this effect.

The comparisons of central interest concern the dependency between vowel length and fricative voicing. Figure 5 shows fricative accuracy broken down by the interaction of gating, fricative identity and inherent vowel duration (Peterson & Lehiste, 1960). The effect of vowel length was assessed by examining the linear interaction of vowel duration with the other manipulations. A significant interaction of vowel duration and gating was observed; $F_1(1,22) = 5.8$, $p < .025$, $F_2(1,9) = 7.3$, $p < .025$. Additionally, the three-way interaction of vowel duration, gating and fricative voicing was significant; $F_1(1,22) = 13.0$, $p < .005$, $F_2(1,9) = 13.0$, $p < .01$. Thus, significance tests bear out the pattern evident in Figure 5. When gated, syllables ending in /s/ are recognized less accurately the greater the inherent duration of the vowel. This suggests that listeners are undercompensating for the effect of inherent vowel duration when interpreting vowel duration as a cue to voicing in syllable-final /s/. In contrast, accuracy in recognizing /s/ does not depend much if at all on inherent vowel duration when the syllables are heard in sentential context.

## General Discussion

This study showed that in certain cases, listeners misinterpret a durational cue that helps convey the identity of adjacent phonetic segments. Understanding the exact nature of this misinterpretation and the circumstances in which it occurs presents an interesting conceptual challenge. Accuracy in recognizing /s/ depended on inherent vowel duration when syllables were gated from sentential context, but improved considerably and was not dependent on inherent vowel duration when syllables were heard in sentential context. The curious nature of this pattern is made clear by comparing it to two other more straightforward kinds of contextual benefit that derived from hearing the syllables in sentential context. The presence of sentential context caused greater improvement in recognizing fricative voicing when the effect of phrase position on vowel duration was incongruent with voicing than when it was congruent. Presumably, hearing the syllable in context allowed listeners to factor out the influence of phrase position on vowel duration and therefore interpret vowel duration more accurately as a voicing cue (Gordon, 1989). The presence of

context also caused greater improvement in recognizing fricative voicing when the voicing of the immediately succeeding consonant was incongruent with the voicing of the fricative. Presumably, hearing the syllable in context allowed listeners to factor out any effects that the voicing of the following consonant had on the acoustic cues to voicing in the fricative. In both these cases, the sentential context provided information -- either phrase position or voicing of the following consonant -- that was of direct relevance to the form of the acoustic cues to voicing in the fricative. In contrast, the elimination of the dependence of accuracy in recognizing /s/ on inherent vowel duration could not have resulted from such a direct provision of relevant information by the sentence context. The gated syllables already provided sufficient acoustic information to identify the vowels whose inherent durations were being inappropriately attributed to the final /s/.

One conceivable way in which sentential context could indirectly eliminate the dependency between accuracy in recognizing /s/ and the identity of the preceding vowel would build on its direct effects in overcoming other impediments to recognizing the gated syllables. On this account, listeners would compensate for the effect of vowel identity on vowel duration before interpreting it as a cue to voicing of the final fricative. Compensation would not be perfect due to variability in the relation between vowel identity and vowel duration. Ordinarily, imperfect compensation would not present a problem because of the differences in the distributions of vowel durations for voiced and voiceless consonants, and because there are other cues to voicing besides vowel duration. However, when the syllables are gated, acoustic support for identifying the fricative is weakened because of the loss of relevant information such as the voicing of the following consonant and the effects of phrase position on vowel duration. Thus, for gated syllables the imperfect nature of the compensation for the effect of vowel identity on vowel duration is enough to tip the balance given the other impediments to recognition. This model, however, can not be correct because it predicts that the interaction between gating and vowel duration ought to further interact with phrase position and/or voicing of the following consonant. These interactions were not observed. Decreased accuracy in recognizing /s/ due to the absence of context depended on inherent vowel duration even

when phrase position and voicing of the following consonant were congruent with the fricative (i.e., phrase-internal position and a following consonant of /t/). In this situation, the decrement due to gating was -0.8% for /ɪ/, 2.5% for /ɛ/, 9.2% for /i/ and 9.6% for /e/. Here, imperfect compensation for the effects of vowel identity on vowel duration could not be seen as tipping a balance against perceiving /s/ produced by gating, because gating had shifted the balance in favor of perceiving /s/.

Of course, the way it is was originally suspected that sentential context might reduce the dependency of fricative accuracy on vowel length was that it might lead to more accurate identification of the vowel and therefore to better compensation for its inherent duration. As it turned out, accuracy in recognizing vowels was so high in the gated condition (96.4%) that there was little room for improvement in the non-gated condition (97.4%), rendering that account untenable. However, there is another way in which the presence of sentential context might directly affect the recognition of adjacent vowel-consonant pairs cued in part by a common acoustic dimension. Undercompensation for inherent vowel duration with the gated syllables may have resulted from a disruption of the normal order in which the vowel and consonant are recognized. The idea that listeners compensate for vowel identity when interpreting vowel duration as a cue to fricative voicing implies that vowel recognition at least partially precedes fricative recognition. One consequence of gating the syllables is that listeners are deprived of a basis, the initial part of the sentence, for predicting the location of the syllable in time. This may delay phonetic processing enough so that acoustic information for recognizing all of the segments in the syllable is available before identification begins. This could change a natural left-to-right sequence of recognizing the vowel and consonant into a more simultaneous process. If this were the case, then compensation for the effect of vowel identity on vowel duration would not occur (or would occur less effectively) and some of the effects of vowel identity on vowel duration would be inappropriately attributed to fricative voicing. A second, less specific, acount would relate the dependence of /s/ accuracy on vowel duration to a general disruption of processing temporal aspects of the signal that stems from gating. The gated syllables are heard in isolation but do not have the prosody of syllables spoken in isolation. By

violating some of the listeners' expectations about temporal aspects of the syllable, this may create some further difficulty in correctly attributing temporal variation in the signal to an underlying phonetic source. Further study will be needed to determine whether some version of either of these two hypotheses -- disruption in the order of processing segments and general disruption of temporal processing -- provides a good acount of the role of sentence context in promoting appropriate attribution of vowel duration to syllable-final consonant voicing independent of vowel identity.

The results of the two experiments reinforce the idea that accurate perception of a phonetic segment requires that listeners use the extended prosodic context in which the segment was spoken. They supported earlier findings concerning listeners' use of prosodic information to factor out extra-syllabic sources of variation in temporal cues to segment identity. In addition, they demonstrated that extended prosodic context plays an important role in enabling listeners to accurately interpret intra-syllabic effects of adjacent phonetic segments on a common durational cue.

## ACKNOWLEDGMENT

The research reported in this paper was supported by grant 80-0416 to Harvard University (Peter C. Gordon, Principal Investigator) from the Air Force Office of Scientific Research Life Sciences Directorate.

## References

Ainsworth, W.A. (1972). Duration as a cue in the recognition of synthetic vowels. ***Journal of the Acoustical Society of America, 51***, 648-651.

Denes, P. (1955). Effect of duration on the perception of voicing. ***Journal of the Acoustical Society of America,*** 27, 761-764.

Fowler, C.A. (1980). Coarticulation and theories of extrinsic timing. ***Journal of Phonetics,8***, 113-133.

Gordon, P.C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. ***Perception & Psychophysics, 43,*** 137-146.

Gordon, P.C. (1989). Context effects in recognizing syllable-final /z/ and /s/ in different phrasal positions. ***Journal of the Acoustical Society of America, 86***, 1698-1707.

Klatt, D.H. (1976). Linguistic use of segmental duration in English: Acoustic and perceptual evidence. ***Journal of the Acoustical Society of America, 59,*** 1208-1221.

Luce, P.A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. ***Journal of the Acoustical Society of America, 78,*** 1949-1957.

Martin, J.G. (1970). On judging pauses in spontaneous speech. ***Journal of Verbal Learning and Verbal Behavior, 9,*** 75-78.

Miller, J.L. (1981). Effects of speaking rate on segmental distinctions. In P.D. Eimas & J.L. Miller (Eds.), ***Perspectives on the study of speech*** (pp. 39-74). Hillsdale, NJ: Erlbaum.

Miller, J.L. (1987). Rate-dependent processing in speech perception. In A. Ellis (Ed.), ***Progress in the psychology of language (Vol. 3)*** (pp. 119-157). Hillsdale, NJ: Erlbaum.

Peterson, G.E., & Lehiste, I. (1960). Duration of syllable nuclei in English. ***Journal of the Acoustical Society of America, 32,*** 693-703.

Port, R.F., Al-Ani, S., & Maeda, S. (1980). Temporal Compensation and Universal Phonetics. ***Phonetica, 37,*** 235-252.

Soli, S.D. (1982). Structure and duration of vowels together specify fricative voicing. ***Journal of the Acoustical Society of America, 72 (2),*** 366-378.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. ***Journal of Experimental Psychology: Human Perception & Performance, 7,*** 1074-1095.

Verbrugge, R.R., & Shankweiler, D.P. (1977). Prosodic information for vowel identity. ***Journal of the Acoustical Society of America, 61,*** S39 (Abstract).

Table 1. The linguistic materials used to generate the speech stimuli.

Syllables: /tIz/ /tɛz/ /tiz/ /tez/ /tIs/ /tɛs/ /tis/ /tes/

Carrier Sentences:

If Ted read _____, Tom could get upset.
If Ted read _____, Dave could get upset.
If Ted read _____ directly, Dave could get upset.
If Ted read _____ today, Tom could get upset.

Figure Captions

Figure 1. Mean recognition accuracy for fricatives in Experiment 1.

Figure 2. Mean recognition accuracy for vowels in Experiment 1.

Figure 3. Mean recognition accuracy for fricatives in Experiment 2.

Figure 4. Mean recognition accuracy for vowels in Experiment 2.

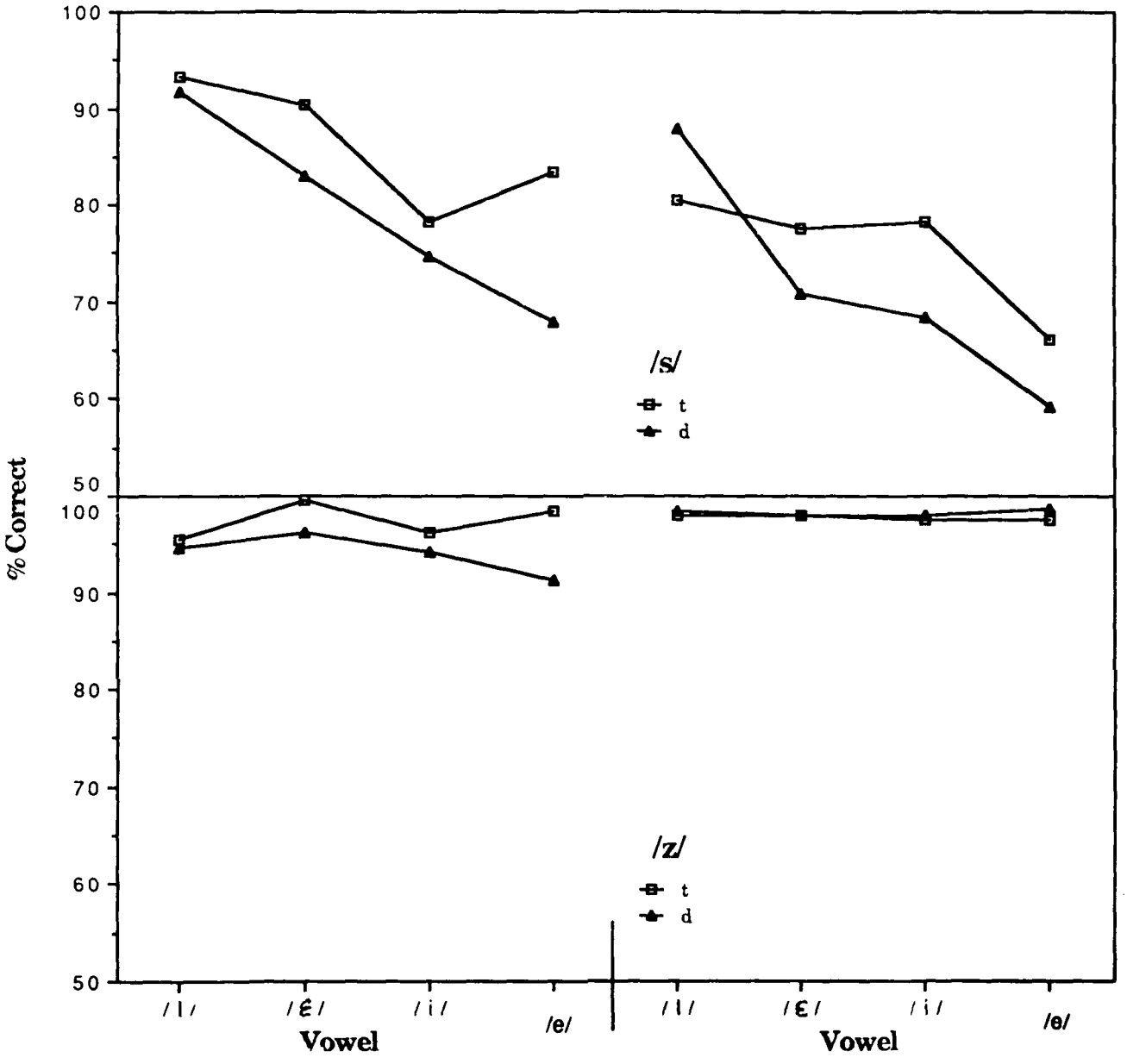
Figure 5. Comparison of mean accuracy for fricatives in Experiments 1 and 2.

**Footnotes**

1. The vowels are as follows: /e/ as in "bait", /i/ as in "beat", /ɛ/ as in "bet", an /ɪ/ as in "bit".

2. It should be noted that the dependency between vowel and fricative recognition was studied by examining how the listeners' identification of a segment related to the intended (or at least instructed) articulations of the speaker. Alternatively, this dependency might have been presented in terms of how a listener's identification of one segment related to his or her identification of the other segment (cf. Mermelstein, 1978). Given listeners' high recognition accuracy for vowels, such a presentation would have yielded a very similar picture. Recognition accuracy for vowels was so high that it did not matter whether performance was broken down by intended fricative or responded fricative -- performance was near ceiling in either case. Conversely, it did not matter whether fricative identification was broken down by intended vowel or response vowel because the two were nearly identical. Analyzing performance relative to intended articulation offered the advantage of a completely balanced design and it also allowed vowel identity and fricative identity to be treated in the same manner as the linguistic manipulations (phrase position and voicing of the following consonant) that listeners did not identify.

# Figure 1.




**Phrase Internal** **Phrase Final**

Figure 2

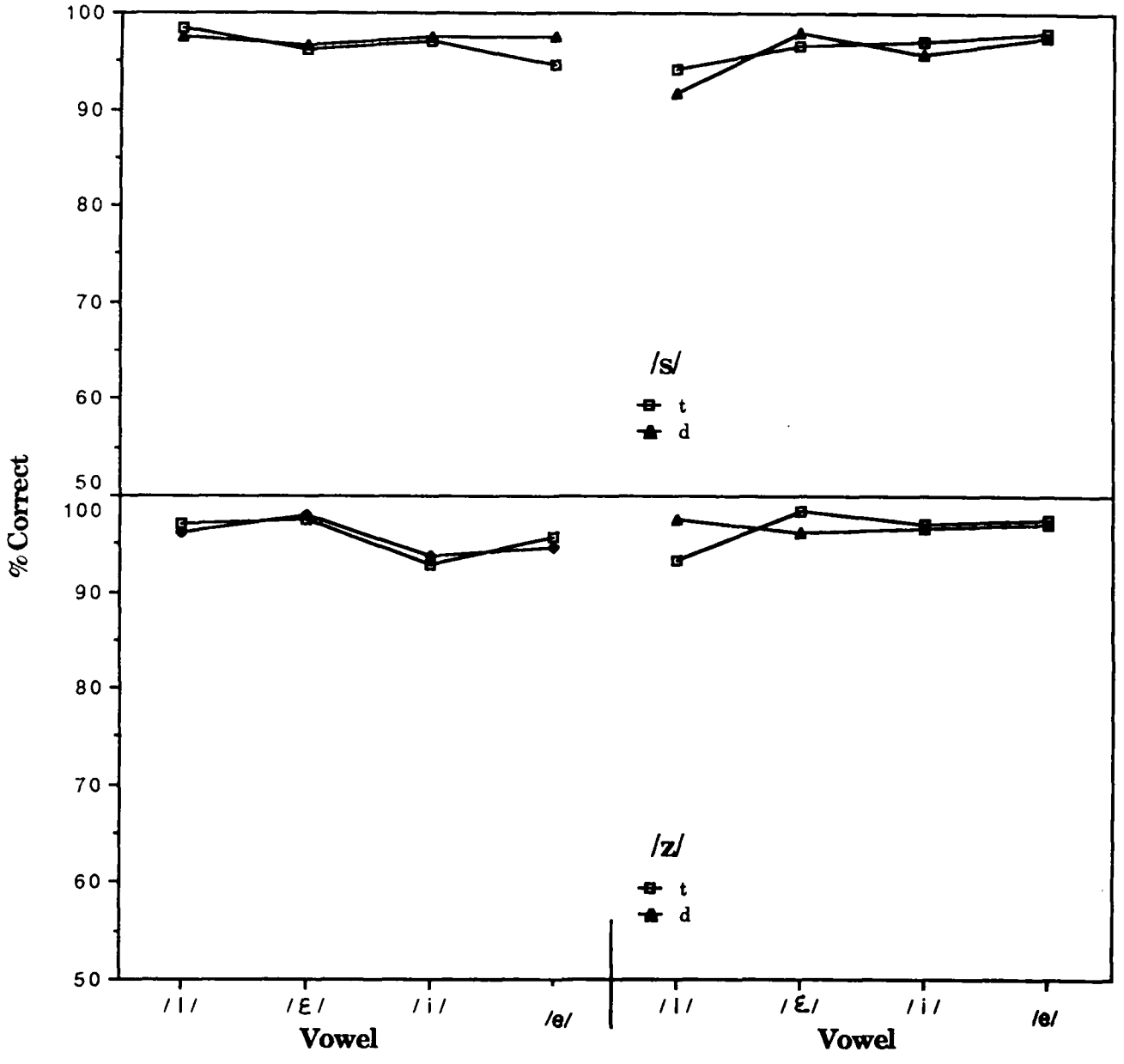
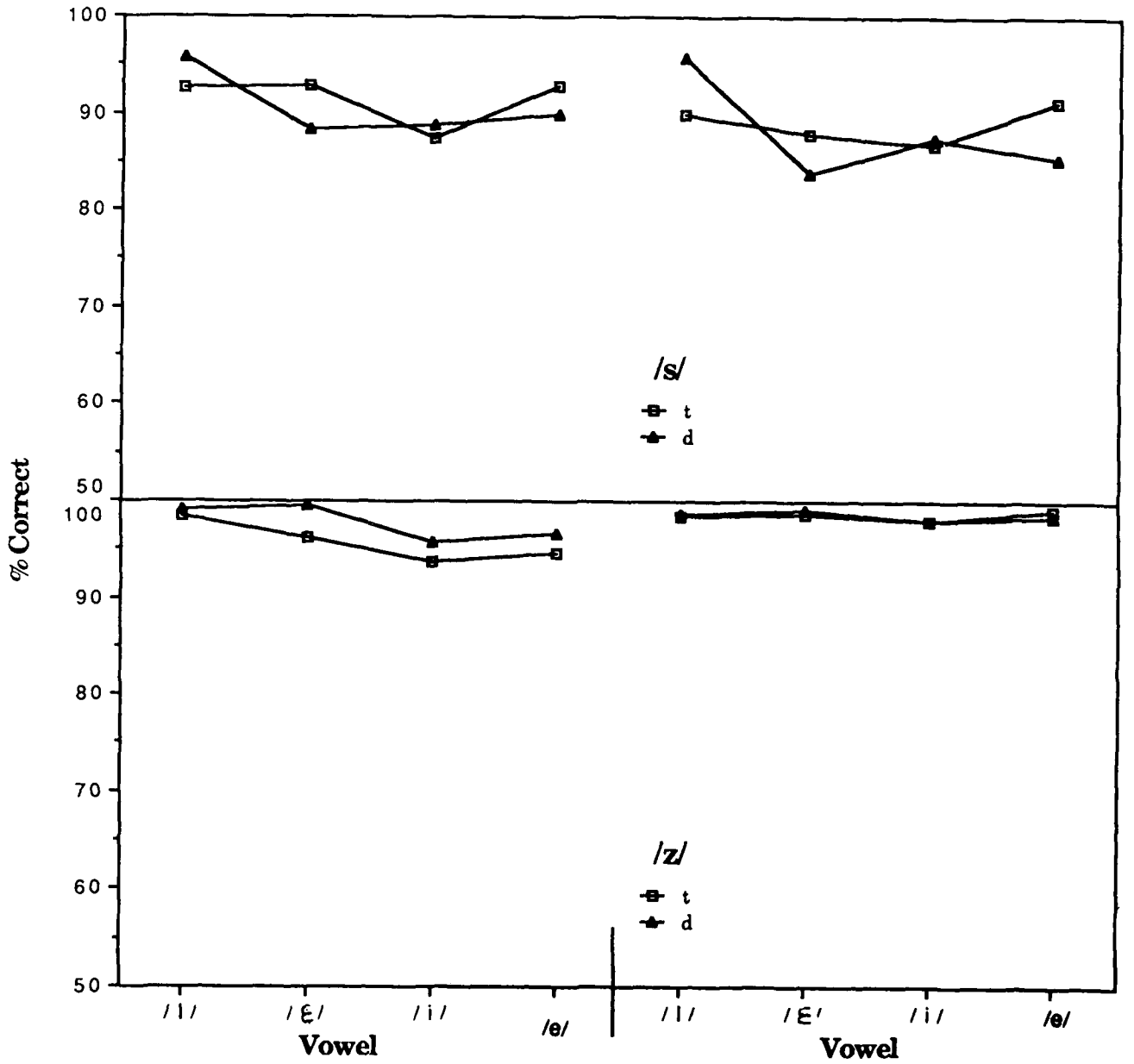
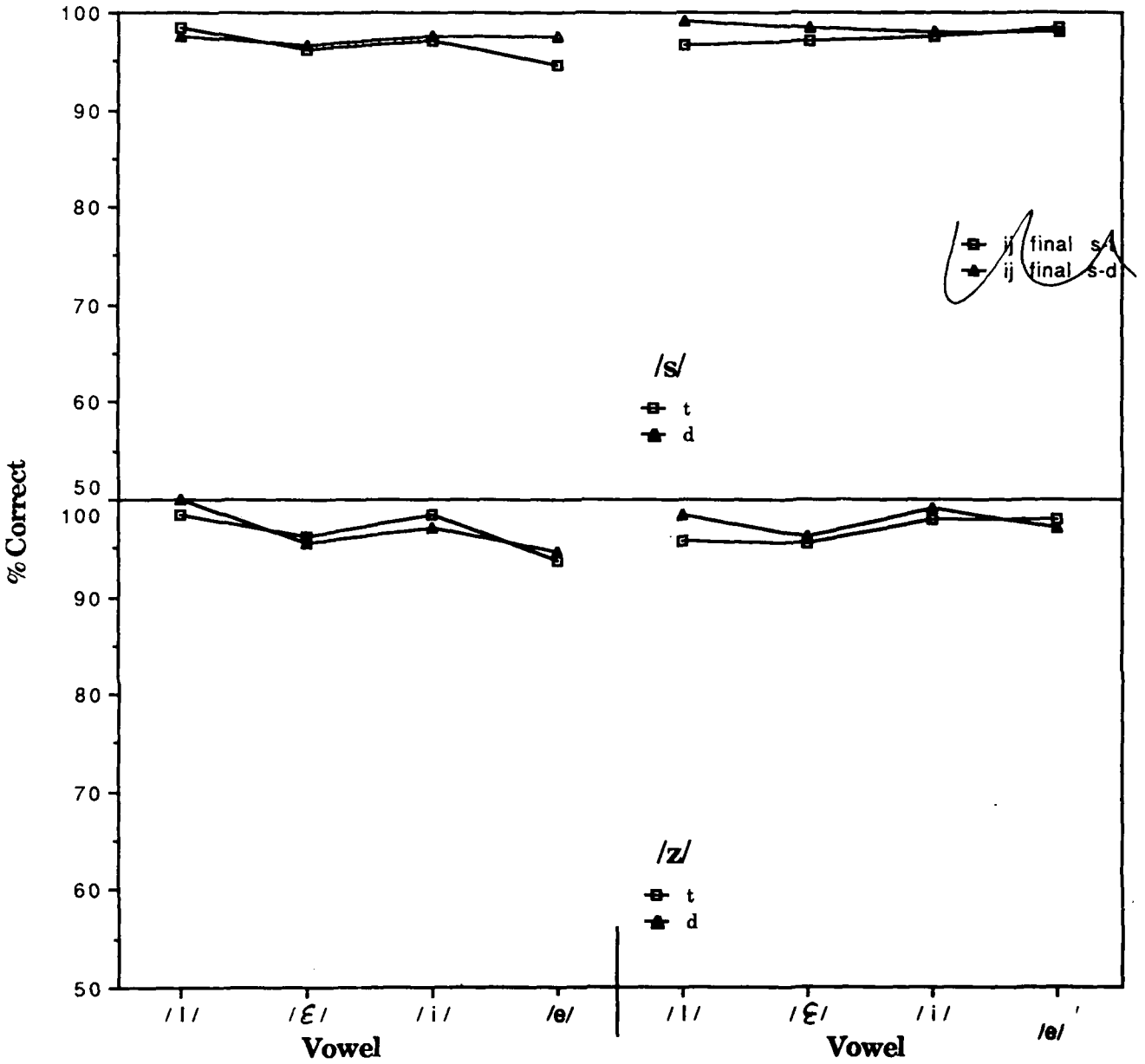

100
90
80
70
60
50
% Correct
/s/
t
d
/z/
t
d
/ɪ/
/ɛ/
/i/
/e/
Vowel
Phrase Internal
Phrase Final

Figure 3

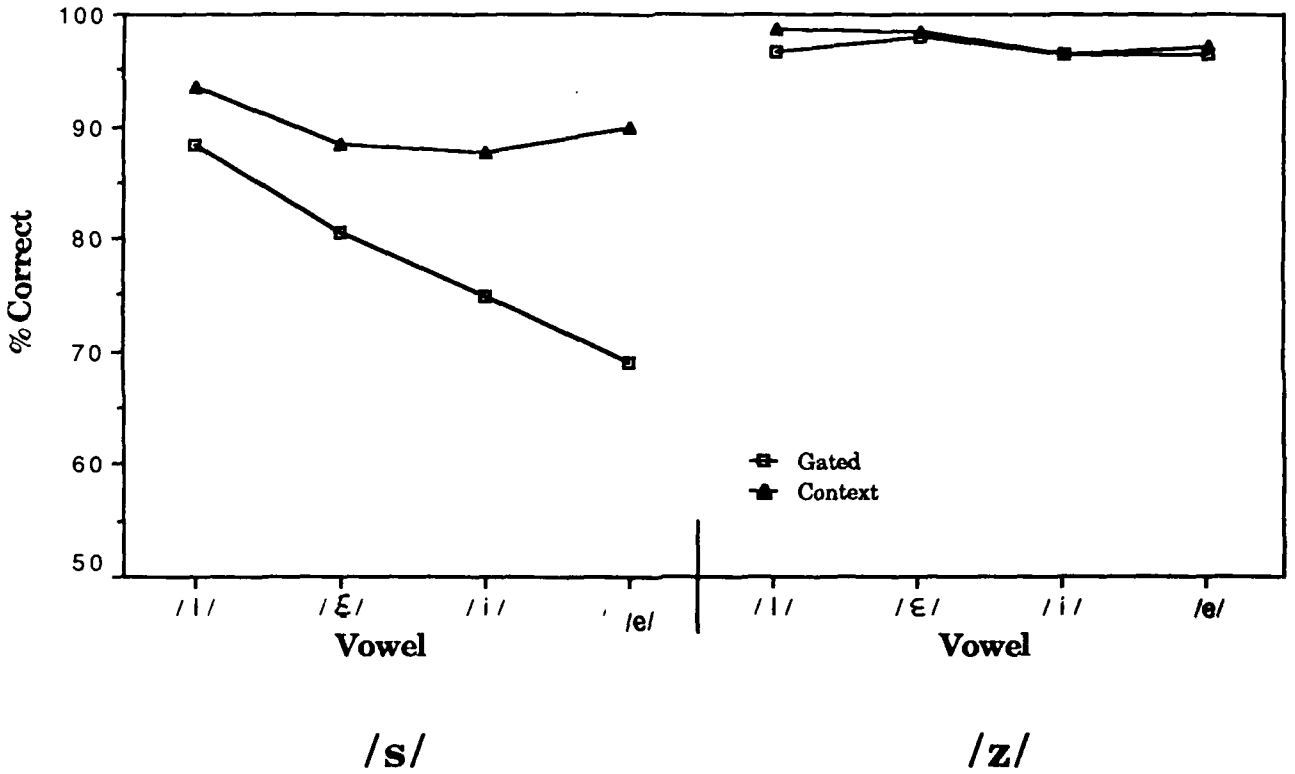



**Phrase Internal** **Phrase Final**

# Figure 4.


100
90
80
70
60
50
% Correct
ij final s-t
ij final s-d
/s/
t
d
/z/
/ɪ/
/ɛ/
/i/
/e/
Vowel
Phrase Internal
Phrase Final

# Figure 5.


100
90
80
70
60
50
% Correct
Gated
Context
/ɪ/
/ɛ/
/i/
/e/
Vowel
/ɪ/
/ɛ/
/i/
/e/
Vowel

/s/ /z/

# Coming to Terms with Stress:
# Effects of Stress Location in Sentence Processing

David W. Gow, Jr. Peter C. Gordon

Harvard University

Running Head: Effects of Stress Location

Send page proofs to:

Peter C. Gordon
Department of Psychology
Harvard University
33 Kirkland St.
Cambridge, MA 02138

**Abstract**

The purpose of this research was to determine the role of syllabic stress in language processing as it changes from the early on-line processing of speech to the later representation of a sentence in memory. Experiment 1 used a syllable monitoring task while Experiment 2 used a probe task in which subjects heard a sentence and then were asked to determine whether a probe syllable had occurred in the sentence. In the monitoring task, stressed syllables were detected more rapidly in word-initial position but unstressed syllables were detected more rapidly in word-final position. This is interpreted as evidence that lexical stress is used on-line to guide lexical access and/or lexical segmentation. In the probe task, stress facilitation occurred in both positions. This suggests that stress is independently represented in the post-perceptual memory of a sentence. The probe task may therefore be valuable as an implicit measure of lexical stress.

The role of stress in the processing of spoken language has been the subject of considerable experimentation and theorizing. Stress effects have been found in a wide variety of on-line comprehension tasks including phoneme monitoring (Shields, McHugh & Martin, 1974; Cutler, 1976), mispronunciation monitoring (Cole & Jakimik, 1978, 1980), embedded-word monitoring (Cutler & Norris, 1988) and shadowing (Small & Bond, 1982), as well as memory-dependent phenomena such as the tip-of-the-tongue state (Brown & McNeill, 1966) and malapropisms (Fay & Cutler, 1977). Theoretical accounts of these effects have variously postulated roles for stress in lexical access (Cutler, 1976; Cutler & Norris, 1988; Bradley, 1980; Grosjean & Gee. 1987), the anticipatory allocation of attention (Shields. et al., 1974; Pitt and Samuel, in press; Meltzer et al., 1976), and perceptual encoding (Lieberman, 1965). In this paper we examine changes in the effect of stress on syllable accessibility between the early on-line phases of comprehension and the representation of a sentence in short-term memory. This is done by comparing the effect of syllable stress in a traditional monitoring procedure with its effect in a new procedure in which listeners are probed for syllables after they have heard an entire sentence. The monitoring task shows that the level of stress interacts with the position of the stressed syllable within a word, providing new evidence in support of the idea that lexical access during continuous speech is expedited by the occurrence of stressed syllables. In contrast, the probe task shows that stress speeds memory retrieval regardless of position in a word. This indicates that stress information is retained in the representation of a sentence independently of other factors with which it interacts in the on-line processing of a sentence. Retrieval time from memory may thus provide a good implicit measure of stress that could supplement phonological intuitions.

## Roles of Stress in On-line Sentence Comprehension

Results from monitoring tasks have been the principle inspiration for theorizing about the role of stress in on-line sentence comprehension. Broadly considered, there have been three types of explanations of the better performance observed on stressed syllables: (1) stressed

syllables are acoustically more salient than unstressed syllables, (2) listeners are able to anticipate the location of stressed syllables, and (3) lexical access from continuous speech depends on stressed syllables in such a way as to facilitate their processing.

The simplest explanation is that the acoustic characteristics of stressed syllables make them easier to perceive than unstressed syllables. While there is no simple mapping between the acoustic characteristics of syllables and their perceived stress, it is generally the case that stressed syllables have greater duration and amplitude than unstressed syllables (Fry, 1958). Several studies support the view that stress facilitates perception. Bond and Garnes (1980) found that stressed syllables are very rarely misperceived in fluent speech. Similarly, Kozhevnikov and Chistovich (1965) found that stressed syllables are detected more consistently than unstressed syllables in noisy environments, and Lieberman (1965) found that the same result holds in the recognition of words excised from fluent speech. It makes sense that speakers would try to articulate stressed syllables in a way that would make them easily identifiable because such units bear a heavy communicative burden; Huttenlocher and Zue (1984) found that stressed syllables in English convey significantly more distinctive lexical information than unstressed syllables. Interestingly, simple intelligibility cannot explain the full range of stress effects found in on-line sentence processing. Shields et al. (1974) found that lexical stress facilitation effects in phoneme monitoring do not depend entirely on the acoustic characteristics of the target-bearing word and Cutler (1976) obtained similar results in an experiment in which contrastive stress was manipulated. These studies demonstrated that stressed targets are facilitated more in their natural sentential contexts than when they are spliced into the positions of unstressed syllables. This suggests that at least part of the role played by stress in processing is a matter of the anticipation of stressed syllables.

A second explanation of the role of stress in sentence comprehension is based on this idea. It has been suggested that stress-timed languages like English allow anticipation of the location of stressed syllables based on the perceived rhythm of the preceding sentential context

(Martin, 1972; Meltzer et al., 1976). Listeners may use these expectations to focus attention on upcoming stressed syllables, thus making efficient use of limited attentional capacities. According to one version of this explanation (Meltzer et al., 1976), listeners are able to anticipate the precise temporal location of the stressed syllable. Support for this view was obtained by showing that phoneme-monitoring times are slowed when targets are temporally displaced by inserting brief periods of silence between words. However, Mens and Povel (1986) have argued that this result is attributable to phonetic discontinuities in Meltzer et al.'s edited stimuli. When Mens and Povel created temporal displacements using a method which did not introduce phonetic discontinuities, they found no slowing of reaction times for displaced targets. This finding undermines the results of Meltzer et al., but it does not necessarily disprove their thesis or some version of it.

One possibility is that expectancies operate on a higher level of representation than actual physical time. In recent years, linguists have developed a number of metrical theories that characterize the representation of stressed and unstressed syllables as consisting of a hierarchical organization with regular patterns known as feet (Liberman, 1975; Liberman & Prince, 1977; Hayes, 1981; Selkirk, 1980). These representations do not include the kind of precise temporal information required by Martin's (1972) theory, but they may capture a regular stress pattern that could be exploited in processing. Liberman and Prince (1977) and Buxton (1983) point out an obvious limitation of using this representation to derive expectancies during on-line processing. A listener would not be able to recover such a hierarchical representation before hearing the whole sentence. However, it is possible that listeners use knowledge of metrical principles to constrain their hypotheses about the location of stressed syllables. A central principle of metrical phonology is that representations of stress are adjusted to avoid "stress clash", the juxtaposition of stressed syllables (Liberman, 1975). Knowledge of this rule does not allow listeners to predict the location of stressed syllables with great accuracy, but it does suggest an

advantage to limiting the amount of attention directed to syllables immediately following stressed syllables.

The third type of explanation asserts that syllable stress affects lexical access in such a way as to facilitate the processing of stressed syllables. According to one version of this hypothesis, advanced by Cutler (1976) and by Bradley (1980), lexical access begins with the identification of a stressed syllable which activates a cohort of words which contain it. This cohort is then pared down to a single lexical entry on the basis of the surrounding syllables. The process facilitates monitoring of stressed syllables and their constituent segments because word recognition initiated by the stressed syllable provides top-down information that allows rapid verification of a target unit's identity. Evidence from Foss and Swinney (1973) and Segui, Frauenfelder and Mehler (1981) suggests that top-down lexical information facilitates phoneme and syllable monitoring. Cutler's (1976) account was originated to explain the stress facilitation found in phoneme monitoring. It is also supported in a general way by results showing that stress patterns are readily retrievable from the lexicon even when segmental information may be difficult to access (Brown & McNeill, 1966; Fay & Cutler, 1977). Other evidence suggests that even if stressed syllables play a special role in lexical access, it is not in the initial activation of a cohort. Cutler (1986) performed a cross-modal lexical priming task using non-morphological homophones with differing stress patterns as primes. She found that both meanings associated with the homophone provided facilitation for semantically related words in a lexical decision task when the probe was presented immediately after the prime. If words were organized in the lexicon on the basis of stressed syllables, then the two meanings should be associated with lexical entries that belong to different stress cohorts. The lack of a distinction between the meanings in the immediate lexical decision provides strong evidence that this is not the case. It is possible though that stress plays a role in the disambiguation of such homophones after initial activation of a homophone pair. When the lexical-decision probe was delayed 250 msec facilitation was limited to targets semantically related to the particular homophone that served as

a prime. However, because the homophones were presented in sentences, it is possible that this disambiguation was achieved via sentential context rather than lexical stress.

According to a second version of the lexical access hypothesis (Cutler & Norris, 1988), stressed syllables trigger word segmentation in speech perception, with listeners tentatively placing word boundaries in front of each stressed syllable that they hear. This theory is supported with evidence from an embedded-word monitoring task in which subjects hear bisyllabic non-words that contain stressed words. Cutler and Norris (1988) found that subjects detected target words faster when the non-word consisted of a stressed syllable followed by an unstressed syllable, than when it consisted of two stressed syllables. They argued that word detection is interfered with in the latter condition by a division of the target word induced by the stressed second syllable of the nonsense word. This interpretation of the role of stress in on-line sentence processing can also account for the phoneme monitoring data which has been marshalled in support of the lexical access hypotheses since lexical segmentation must precede lexical access in continuous speech. Thus, monitoring effects which seem to reflect lexical access may actually reflect the combination of lexical access and lexical segmentation.

Clearly the three classes of explanation for on-line stress effects are not mutually exclusive, nor does any one of them appear uniquely strong. The acoustic-salience explanation does not account for results obtained in studies that have matched the acoustic properties of stressed and unstressed targets, yet it seems very likely that acoustic salience plays some role in the on-line processing of stress. An explanation that makes use of of anticipatory processing seems necessary to account for stress-facilitation effects that depend on pre-target patterns (Shields et al., 1974; Cutler, 1976; Pitt & Samuel, in press), though anticipatory processing does not readily account for lexically-mediated stress effects. Conversely, the lexical-access explanations do not handle anticipatory effects, nor do they account for stress effects observed with non-word targets (Shields, et al., 1974). All three of these mechanisms have empirical

support and theoretical appeal. It seems likely that stress influences a variety of processes during the on-line comprehension of spoken language.

Current Goals

One goal of the present work is to further clarify the role of syllable stress in the on-line processing of spoken language. Experiment 1 does this by examining possible dependencies among syllable stress, syntactic predictability of stress, and position of stress within a word. A second goal is to explore the relationship between the role of stress in early on-line comprehension and in the later post-perceptual representation of a sentence. Experiment 2 makes this possible by using the same stimuli as Experiment 1, but having subjects respond to a probe syllable after having heard a sentence, rather than monitoring for the syllable while listening to the sentence. Comparing the results of the two experiments allows us to begin to chart the impact of stress from the initial encoding of speech to the memory representations that likely participate in the higher-level comprehension of language. In addition to providing information about the processing of stress, consideration of these post-perceptual representations has potential bearing on structural issues. As has often been noted (e.g., Swinney, 1984), psychologists have typically focused their efforts on experimental analyses of the processing of language while linguists have typically worked with intuitions about its post-perceptual structure. The probe task used in Experiment 2 potentially offers an implicit measure of syllable stress in post-perceptual representations that may be useful in cases where intuitive judgments and acoustic measurements, by phonologists and phoneticians, have yielded conflicting results (cf. Cooper & Eady, 1986; Liberman & Prince, 1977).

## Experiment 1

This experiment uses a syllable-monitoring task to study how stress level interacts with position of stress within a word and with the syntactic predictability of stress location. Previous

monitoring research on stress has focused on the detection of word-initial targets, even though syllable stress often occurs in non-initial position. In this experiment, a controlled manipulation of stress level in different word positions is obtained by comparing monitoring times to first and second syllables in bisyllabic noun-verb homophone pairs such as CONduct and conDUCT (where capitalization indicates the placement of primary stress). For such homophones, nouns have stress on the first syllable while verbs have stress on the second syllable, thereby allowing us to examine stress effects in word-initial and word-final positions. While most previous stress research using monitoring has only used initial targets, other research has addressed target detection as a function of position within a word. Marslen-Wilson (1984) used one to three syllable words and non-words in a monitoring task and varied the position of the target syllable in each word. He found a serial position effect, with early syllables detected more slowly than later syllables. Segui and Frauenfelder (1986) and Pitt and Samuel (in press) supply converging results by showing a reaction time advantage for the detection of word-medial versus word-initial phonemes. These findings are consistent with Marslen-Wilson's cohort model as well as other models in which lexical access is initiated on the basis of the segmental identity of the initial portion of a word. The current examination of the joint effects of position and stress has potential implication for the cohort-type model as well as for the lexical-interaction accounts of stress discussed earlier. The Marslen-Wilson and Tyler (1980) cohort model works from left to right in activating and narrowing down the cohort, making no provision for the use of stress. The lexical-interaction models of stress processing have been formulated to account for results obtained with stress in word-initial position. The generality of both the stress-cohort model (Cutler, 1976; Bradley, 1980) and the stress-segmentation model (Cutler & Norris, 1988) is limited by their failure to consider possible dependencies of stress effects on lexical position.

Manipulation of the syntactic ambiguity of the target word was achieved by using unambiguous initial sentence fragments that strongly constrained the word class (and therefore the stress position) of the words containing the target syllable, as well as ambiguous initial

sentence fragments that were easily followed by either version of the homophone. By varying both the syntactic predictability and the stress of the target syllables, this experiment allows us to see whether the magnitude of stress facilitation depends on whether the location of a stressed syllable is predictable based on higher-level representations of a sentence as might be expected under some versions of stress anticipation models.

Method

*Subjects*. Thirty two Harvard University students were paid $5.00 to participate in a single session lasting 40 minutes. Half of the subjects were male and half were female. All were native speakers of American English with no known auditory or (uncorrected) visual deficits.

*Stimuli*. Twenty-four bisyllabic noun-verb homophones with a clear stress distinction were used in the study. This word class distinction provided a way of manipulating syllable-position of the stress: nouns had stress on the first syllable while verbs had stress on the second syllable. Each word occurred in four sentences created by combining word class with syntactic ambiguity of the pre-target sentence as illustrated in Table 1. In the ambiguous condition, the words in the initial sentence fragment could be easily followed by either the noun or verb forms of the homophone. Two uambiguous initial sentence fragments were constructed for each homophone, one which strongly constrained the following word to be a noun and the other which constrained it to be a verb. The twenty-four homophones in their four different sentence frames are given in Appendix 1. In addition to the experimental sentences, 96 filler sentences were also constructed.

/ insert Table 1 about here /

All of the sentences were spoken aloud by an adult male speaker of American English. The speaker was asked to read each sentence to himself before reading it aloud in order to ensure normal intonation. The sentences were spoken in a sound-attenuating chamber and recorded on

a cassette deck using a Shure SM59 microphone. The recordings were low-pass filtered at 4.7 kHz, digitized at 10kHz, and equated for amplitude. The digitized sentences were then used as stimuli in the study.

*Procedure.* Subjects were seated at a desk in a sound-attenuating chamber with a CRT placed at eye level roughly 30 inches away. They initiated a trial by pressing a mouse button which caused the target syllable to appear on the screen in conventional spelling. Subjects were instructed to read the target syllable to themselves and think of its sound. After four seconds, a sentence was presented over the headphones. Subjects were instructed to listen to the sentence carefully, and press the left button on the mouse as quickly as possible when they heard the target syllable. Response times were measured from the onset of the target syllable as determined using a waveform editor. The presentation of the audio stimulus was halted when the subject responded. Ten percent of the trials (all of them with filler sentences) did not contain the indicated target syllable in order to minimize disproportional vigilance at the end of the sentence. On negative trials, subjects were not to make any response until they saw the prompt to initiate the next trial.

*Design.* There were eight experimental conditions determined by the combination of target position (initial vs. final), target stress (stressed vs unstressed) and ambiguity of the initial sentence fragment. An individual subject saw a given homophone in only one of the eight conditions, and across subjects each homophone participated equally in all conditions. The 24 experimental trials plus the 96 filler trials gave a total of 120 trials which were grouped into six blocks. The first three blocks contained only filler trials to allow performance in the task to stabilize.

## Results

Figure 1 shows the mean response times for the experimental conditions. Response times shorter than 500 msec and longer than 2000 msec were considered outliers and were excluded

from the analyses (cf. Cutler, 1976). Separate analyses of variance were computed by subjects ($F_1$) and items ($F_2$). There were two significant main effects. Mean response times for word-final targets were faster than for word-initial targets: $F_1(1,31) = 171.3$, $p < .001$; $F_2(1,23) = 112.0$, $p < .001$. Mean response times were also faster for unambiguous targets than for ambiguous targets: $F_1(1,31) = 10.9$, $p < .005$; $F_2(1,23) = 4.9$, $p < .05$. There was no main effect of stress: $F_1(1,31) < 1$; $F_2(1,23) < 1$. However, stress did interact significantly with position of the target syllable within the target-bearing word: $F_1(1,31) = 11.5$, $p < .005$; $F_2(1,23) = 11.2$, $p < .005$. Stressed syllables were detected faster than unstressed syllables when they were word-initial ($t_1(31) = 2.48$, $p < .05$; $t_2(23) = 2.54$, $p < .05$), but were detected more slowly than unstressed syllables when they were word-final ($t_1(31) = 2.38$, $p < .05$; $t_2(23) = 2.54$, $p < .05$). This was the only significant interaction.

Missed targets were infrequent, averaging three percent over all the trials and less than one percent for the experimental trials (which took place after subjects had performed three practice blocks).

## Discussion

The pattern of results confirms some previous findings as well as offering new evidence about the role of stress in sentence comprehension. The finding that subjects are slower to respond to targets in ambiguous as compared to unambiguous sentences is consistent with previous results indicating that monitoring tasks are sensitive to processing load (Cutler & Norris, 1979; Foss, 1969). The faster response to word-final syllables as compared to word-initial syllables is consistent with results showing that monitoring performance improves over the course of a word (Marslen-Wilson, 1984). With regard to stress facilitation, previous studies using word-initial stress had observed that stressed syllables are detected more rapidly than unstressed syllables. The current experiment replicated this effect for word-initial syllables, but showed that the opposite effect obtains for the second syllable in the bisyllabic words presently

under investigation. This finding, in conjunction with the other findings in the experiment, has bearing on the acoustic-salience, anticipation, and lexical-interaction models of the role of stress in sentence processing.

According to the acoustic-salience account, stress facilitation is observed because the acoustic characteristics of stressed syllables (e.g., relatively greater amplitude and duration) make them easier to detect. All previous research that has ruled out acoustic salience as the sole basis for stress-facilitation has used splicing techniques (Cutler, 1976; Shields et al., 1974) that are a potential source of methodological difficulties. The current study showed that in word-final position, unstressed syllables are detected more rapidly than stressed syllables. This indicates that the acoustic properties of stressed syllables do not always afford a processing advantage over matched unstressed syllables, thereby corroborating the inferences that have previously been drawn using splicing techniques.

According to anticipation accounts of stress-facilitation, listeners use the pre-target speech to determine where to expect stressed syllables and devote more processing resources to those locations. Since the present experiment used sentences with normal prosody, word-final stress should have been just as predictable as word-initial stress. Therefore, anticipation models would predict that stressed syllables should be detected faster than unstressed syllables regardless of their position within a word. The stress-by-position interaction shown in Figure 1 indicates clearly that this is not the case.

As noted above, some researchers have suggested that expectations are generated about the exact moment at which a stressed syllable will occur (Meltzer, et al., 1976), while we have argued that expectations might be based on higher-level representations. The present syntactic ambiguity manipulation allowed us to compare performance when word class (and hence stress location) was strongly constrained by higher-level representations with performance when word class was not constrained. While the main effect of ambiguity

shows that this manipulation affected performance, the absence of an interaction between ambiguity and stress provides no evidence that syntactic predictability led to expectations about stress location that facilitated processing. The effect of ambiguity may be attributable to the syntactic or morphological ambiguity of the preceeding context rather than any ambiguity in the syntactic category of the target word. This interpretation is consistent with work summarized by Cutler and Norris (1979) showing that monitoring times are affected by the processing demands of words immediately preceeding targets. The present results therefore fail to provide any new support for anticipation based accounts of stress facilitation.

Lexical-interaction accounts of stress facilitation receive the most support from the current results, although neither the stressed-cohort model or the stress-segmentation model account for the data without some elaboration. The stressed-cohort model (Cutler, 1976) would predict a clear advantage for words beginning with a stressed syllable, since lexical activation is initiated by the occurence of a stressed syllable. This would explain the stress effects observed in both word-initial and final positions. Stress faciliation occurs in initial position because the stressed syllables initiate lexical access. Because the words with unstressed syllables in second position all had stressed syllables in first position, the apparent facilitation of unstressed syllables in second position is attributable to the lexical access initiated by early stress. This model does not, however, account for the faciliation effect observed for second syllables as compared to first syllables regardless of stress location. Lexical access apparently proceeds on additional grounds besides stressed syllables.

The stress-based lexical segmentation model (Cutler & Norris, 1988) can also account for the pattern of stress effects. If lexical segmentation is triggered by stressed syllables then first syllable stress should lead to the appropriate segmentation of words in the speech stream, and initiate efficient lexical access resulting in facilitation in recognizing stressed over unstressed syllables in initial position. Conversely, stressed syllables in second position should cause inappropriate segmentation, treating the second syllable of a target-bearing word as the first

syllable of a new word. As this would generally not match any entries in the lexicon, it would delay lexical access until the segmentation problem could be corrected through processes guided by non-stress information. The difference in detection latencies for stressed versus unstressed syllables in second syllable position can thus be interpreted as the result of the lexical access advantage gained for target-bearing words due to efficient lexical segmentation triggered by initial stress. Unfortunately, the segmentation account would seem to predict that syllables in words with second-syllable stress would be harder to detect than those in words with first-syllable stress. This did not occur.

While neither of the specific lexical models discussed above can accomodate all of our results without modification, it is clear that the dependence of stress-facilitation on syllable position within a word suggests that lexical processing and stress processing interact strongly in on-line language comprehension.

## Experiment 2

Our focus so far has been on the on the manner in which syllable stress participates in the on-line comprehension of language as measured by syllable accessibility in a monitoring task. In this on-line task, syllable accessibility was found to be influenced by syllable position, syntactic ambiguity and the interaction of syllable stress and position. In Experiment 2, we examine the effect of stress on accessing syllables from the post-perceptual representation of a sentence. The goal of the study is to see whether stress has an independent representation in short-term memory after lexical access and disambiguation have been completed.

There is some evidence that suggests that stress is represented in memory, though this evidence does not bear directly on short-term representations. The frequent use of verse in communities that have an oral tradition for conveying history and culture suggests that patterns of stress facilitate recall. Evidence from the tip-of-the-tongue phenomenon (Brown & McNeill,

1966) and from malapropisms (Fay & Cutler, 1977) suggests that stress information is relatively easy to recall. The present experiment goes beyond this evidence by using speeded responses in a short-term memory probe task to measure syllable accessibility as a function of stress and other factors after a sentence has been heard.

Short-term memory probe tasks have often been used to study syntactic phenomena in spoken language comprehension, and they have proven to be sensitive indicators of the hierarchal structure of sentences. Typically, these tasks have involved auditory presentation of a single sentence utterance followed by visual presentation of one or more probe words. Subjects are asked to determine as quickly as possible whether or not the probe was in the sentence. For example, Suci, Ammon and Gamlin (1967) used two-word probes to demonstrate that noun and verb phrases are psychologically real units of language. Other studies have similarly shown that response time to probes is a valid and sensitive indicator of structural aspects of a previously heard sentence (Caplan, 1972; Walker, 1976), including their phonological form (Green, 1975).

In Experiment 2 we use a probe task to determine whether stress is represented in short-term memory and whether its influence on syllable accessibility depends on its position within a word. If the stress by position interaction found in the first experiment reflects the role of stressed syllables in lexical segmentation or lexical access, then we would not expect stress to interact with position in the memory probe task because these processes should be completed by the time the probe is presented. Similarly, we would expect that the syntactic ambiguity effect found in the monitoring task would not be found in the probe task because ambiguity should be resolved prior to the presentation of the probe.

## Method

*Subjects*. Forty-eight subjects from the same population as the previous experiment served in a single session lasting 45 minutes. None of them had participated in Experiment 1.

They were paid $4.00 plus a bonus of up to $2.00 depending on their speed and accuracy in the task.

*Stimuli, Procedure, & Design.* The sentences were the same as in the previous experiment. However, for half of the sentences (all of which were filler sentences), the probe was changed so that it was not in the sentence. Subjects were tested in the same set-up as Experiment 1. They were instructed to listen to sentences over headphones, and as soon as each sentence ended, a target syllable appeared on the screen, remaining there until the subject pressed a response button on the mouse. Subjects were instructed to press the left button on the mouse if the sentence included the visual probe syllable, and the right button if it did not. Following each of the first five trials, subjects received feedback consisting of their reaction time, response accuracy, and the number of points they earned toward a cash bonus. After these familiarization trials, subjects only received accuracy feedback following incorrect responses. At the end of each twenty-sentence block, subjects received summary feedback including their total number of errors, average reaction time, and the number of points they earned over the course of the block. Trials were arranged so that no experimental sentences appeared in the first three blocks. Eight experimental trials appeared randomly in each of the last three blocks.

## Results

As in Experiment 1, responses with latencies less than 500 msec, or greater than 2000 msec were excluded from the analyses. In addition, after the experiment had been run, it was discovered that one of the homophones, "address", had not been included in all of the experimental conditions, so the data for this word was also excluded from the analyses. Figure 2 shows the mean response times for the various conditions. As in the previous experiment, response times were significantly faster for syllables in word-final position than for syllables in word-initial position; $F_1(1,47) = 16.3$, $p < .005$, $F_2(1,22) = 6.3$, $p < .025$. In contrast to the previous experiment, there was no significant effect of ambiguity; $F_1(1,47) < 1$, $F_2(1,22) < 1$.

However, there was a significant main effect of stress, with responses to stressed syllables being faster than to unstressed syllables; $F_1(1,47) = 4.9$ $p < .05$, $F_2(1,22) = 5.3$, $p < .05$. The effect of stress did not depend on the position of the stressed syllable within the word; $F_1(1,47) < 1$, $F_2(1,22) < 1$. All other interactions failed to reach significance as well.

Figure 2 also shows the average accuracy in each condition. Subjects were more accurate in responding to syllables in word-final position than in word-initial condition; $F_1(1,47) = 6.6$, $p < .025$, $F_2(1,22) = 4.5$, $p < .05$. This was the only significant effect.

Discussion

The results of Experiment 2 show that stress information is included in the post-perceptual representation of a sentence in short-term memory. This inference is based on the finding that probes for stressed syllables were responded to faster than probes for unstressed syllables. In contrast to the results of the monitoring task used in Experiment 1, the stress effect in the current probe task did not depend on the position of stress within a word. This suggests that stress-dependent syllable accessibility in the probe task does not interact with lexical access processes as we hypothesize it does in the monitoring task. A further indication that the probe task taps different processes than the monitoring task is provided by the absence of an ambiguity effect in Experiment 2. This suggests that the on-line syntactic disambiguation processes which affect performance on the monitoring task have been completed by the time that the probe syllable is presented.

The results also showed a significant position effect, with faster responses to word-final as compared to word-initial syllables. When obtained in monitoring tasks, such position effects have been interpreted as reflecting the process of lexical access (e.g., Marslen-Wilson, 1984 and our discussion of Experiment 1). The finding of a significant position effect (both in response time and accuracy) in the post-perceptual probe task could be taken as a challenge to that interpretation. Before this is done, it should be noted that the position effect in the monitoring

task averaged 137 msec while it averaged only 47 msec in the probe task. The larger position effect observed in the monitoring task suggests that it derives at least in part from on-line processes such as lexical access

The most interesting finding of the experiment was that stressed syllables were accessed from memory more rapidly than unstressed syllables, and that this facilitation was independent of stress-position within a word or syntactic ambiguity of the pre-target phrase. The results of the experiment do not inform us of the exact manner in which stress is represented in the memory for a sentence, nor does it inform us of how this representation produces a response-time facilitation in the probe task. It is interesting to note, however, that when confronting the elusive nature of stress some linguists have referred to stressed syllables as being "marked for consciousness" (Selkirk, 1984, p. 10). It appears that one consequence of this marking is ready accessibility from memory.

## General Discussion

Experiments 1 and 2 demonstrate that stress plays a role in the on-line processing of continuous speech as well as the post-perceptual accessability of syllables. We believe that these roles reflect independent mechanisms which capitalize on the special acoustic (Fry, 1958) and distributional (Kelly & Bok, 1988) qualities of stressed syllables in speech.

The stress effects found in the monitoring task used in Experiment 1 suggest that word-initial stressed syllables facilitate the identification of words in continuous speech. Word identification then provides a source of top-down information which aids in the final identification of syllables. Our stress facilitation effects are consistent with the existence of a lexical segmentation mechanism similar to the one proposed by Cutler and Norris (1988) and/or a lexical access system which uses stressed syllables as one means of activating items in the lexicon. However, the reaction time advantage found for both stressed and unstressed word-final

syllables suggests that lexical access may also be initiated prior to the occurrence of a stressed syllable. This could occur either through the existence of several independent lexical access routes, or by a single access process which is facilitated by the distributional and acoustic distinctiveness of stressed syllables.

The contrast between these results and the results of the probe task used in Experiment 2 supports this interpretation. In the monitoring task, stress interacted with word position, while stress facilitation in the probe task was independent of syllable position. Such a difference is to be expected if the stress effects found in Experiment 1 are attributable to uniquely on-line processes such as lexical access and lexical segmentation. These processes should be completed prior to the presentation of the probe in Experiment 2, and therefore would not affect the processes required to perform the probe task. Thus, the results of the probe task support our interpretation of the monitoring task as reflecting on-line processes.

The results of Experiment 1 also shed light on the types of information which are used to predict stress. Previous studies employing editing techniques have shown that disrupting the prosodic structure of utterances significantly reduces the size of stress effects (Shields et al., 1974; Cutler,1976). This result is generally accepted as evidence that prosodic information is used to predict the location of upcoming stressed syllables. The results of Experiment 1 show that if there is any such anticipatory processing, it does not use syntactic/semantic information to inform its predictions. If it did, we should have found an interaction of stress and ambiguity, or a three-way interaction of stress, ambiguity and syllable position. The significant effect for ambiguity that we did find may be attributable to the processing demands required to disambiguate the syntactic/semantic structure of the pretarget context (Foss and Blank, 1977; Cutler and Norris, 1979).

The results of Experiment 2 show that stress is represented independently in the post-perceptual memory of a sentence; a finding that may have important procedural implications for

research on stress. Lexical stress is notoriously difficult to measure. While it is generally accepted that perceived stress is correlated with certain simple acoustic features, there is evidence that these features alone cannot account for the full range of stress distinctions that can be made by a native speaker of a language (Lieberman, 1965, 1967; Chomsky and Halle, 1968). Furthermore, explicit stress judgments of the type routinely employed by phonologists may lack consistency. Cooper and Eady (1986) have shown that stress assignments of trained phoneticians, naive to Liberman and Prince's (1977) influential metrical theory of stress, differ from those of the theory's proponents on certain critical linguistic constructs. This suggest that explicit judgments of stress assignment are subject to observer bias. This conflict is not resolved by acoustical measurements. Perhaps because perceived stress is due to complex relations among several acoustic dimensions that are not well understood, simple acoustic measurements are not sufficient. Though measuring syllable durations on similar linguistic forms, Cooper and Eady (1986) and Rackerd and Fowler (1984) arrived at different interpretations of the effect of stress on speech timing. Given the problems associated with explicit stress judgments and acoustical measurements, we recognize the need for an experimental procedure to measure the subjective experience of stress. Such a measure could be used to empirically examine the processing implications of sophisticated linguistic theories of stress distribution.

The results of our studies suggest that the subjective experience of syllabic stress is more highly and uniquely correlated with facilitation on the probe task than it is with performance on the monitoring task. This offers the hope that the procedure can be adapted to examine the psychological reality of metrical theory, and determine its relevance to issues in speech processing. Metrical theories (Liberman, 1975;, Liberman and Prince, 1977; Hayes, 1981; and Selkirk, 1984) suggest that the distribution of stress is not adequately described by a simple left to right alternation of stressed and unstressed syllables. Rather, they claim that stress is represented hierarchically through the complex interaction of lexical information and phonological rules. By probing subjects at different points after the presentation of a word in

context, one might examine the timecourse, and the types of information relevant, to non-linear stress assignment. This would allow examination of the validity and processing implications of metrical conceptions of stress.

## References

Bond, Z., & Garnes, S. (1980). Misperceptions of fluent speech. in R. Cole (Ed.), ***Perception and production of fluent speech***. Hillsdale: Lawrence Erlbaum.

Bradley, D. (1980). Lexical representations of derivational relation. In M. Aronoff & M.L. Kean (Eds.), ***Junctures***. Saratoga, CA: Anma Libri.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. ***Journal of Verbal Learning and Verbal Behavior, 5***, 325-337.

Buxton, H. (1983). Temporal predictability in the perception of English speech. In A. Cutler & D.R. Ladds (Eds.), ***Prosody: Models and measurements***. Berlin:Springer-Verlag, 111-122.

Caplan, D. (1972). Clause boundaries and recognition latencies for words in sentences. ***Perception & Psychophysics, 97***, 254-257.

Chomsky, N., & Halle, M. (1968). ***The sound pattern of English***. New York: Harper and Row.

Cole, R., & Jakimik, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), ***Strategies of information processing***. NY: Academic.

Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), ***Perception and production of fluent speech***. Hillsdale: Lawrence Erlbaum.

Cooper, G., & Eady, S.J. (1986). Metrical phonology and speech production. ***Journal of Memory and Language, 25***, 369-384.

Cutler, A. (1976). Phoneme monitoring reaction time as a function of preceding intonation contour. ***Perception & Psychophysics, 20***, 55-60.

Cutler, A. (1986). ***Forebear*** is homophone: Lexical prosody does not constrain lexical access. ***Language and Speech, 29***(3), 201-220.

Cutler, A., & Fodor. J.A. (1979). Semantic focus and sentence comprehension. ***Cognition, 7,*** 49-59.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language specific comprehension strategy. ***Nature, 304,*** 159- 160.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in segmentation of French and English. ***Journal of Memory and Language, 25,*** 385-400.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. Cooper & E. Walker (Eds.), ***Sentence processing.*** New York: Lawrence Erlbaum Associates.

Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. ***Journal of Experimental Psychology:Human Perception and Performance, 14,*** 113-121.

Fay, D.A., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. ***Linguistic Inquiry, 8,*** 505-520.

Foss, D.J. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. ***Journal of Verbal Learning and Verbal Behavior, 8,*** 457-462.

Foss, D.J., & Blank. M.A. (1980). Identifying the speech codes. ***Cognitive Psychology, 12,*** 1-31.

Foss, D.J., & Swinney, D.A. (1973). On the psychological reality of the phoneme: Perception, identification and consciousness. ***Journal of Verbal Learning and Verbal Behavior, 12***, 246-257.

Fry, D.B. (1958). Experiments in the perception of stress. ***Language and Speech, 1***, 126-152.

Green, D.W. (1975). The effects of task on the representation of sentences. ***Journal of Verbal Learning and Verbal Behavior, 14***, 275-283.

Grosjean,F., & Gee,J.P. (1987). Prosodic structure and spoken word recognition. ***Cognition, 25***, 135-155.

Hayes, B. (1981). ***A metrical theory of stress rules***. Bloomington: Indiana University Linguistics Club.

Huttenlocher, D.P., & Zue, V.W. (1984). A model of lexical access based on partial phonetic information. ***Proc. ICASSP-*** 84, 26.4.1-4.

Kelly, M.H., & Bok, J.K. (1988). Stress in time. ***Journal of Experimental Psychology: Human Perception and Performance***, 14(3), 389-403.

Kozhevnikov, V., & Chistovich, L. (1965). ***Speech: Articulation and perception***. US Department of Commerce Translation. IPRS 30, 543. Washington, D.C.

Liberman, M.Y. (1975). ***The intonational system of English***, Doctoral Dissertation, MIT, Cambridge, MA.

Liberman, M.Y., & Prince, A. (1977). On stress and linguistic rhythm. ***Linguistic Inquiry, 8***, 249-336.

Lieberman, P. (1965). On the acoustic basis of perception of stress by linguists. ***Word, 21***(1), 40-54.

Lieberman, P. (1967). ***Intonation, perception and language***. Cambridge: MIT Press.

Marslen-Wilson, W.D. (1984). Function and process in spoken word recognition: A tutorial review. In H. Bouma & D.G. Bouwhuis (Eds.), ***Attention and performance X: Control of language processes***. NY: Academic.

Marslen-Wilson, W.D., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. ***Cognition, 8***, 1-71.

Martin, J.G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. ***Psychological Review, 79***, 487-509.

Meltzer, R.H., Martin, J.G., Mills, C.B., Imhoff, D.L.,& Zohar, D. (1976). Anticipatory coarticulation and reaction time to phoneme targets in spontaneous speech. ***Phonetica, 37***, 159-168.

Mens, L.H., & Povel, D.J. (1986). Evidence against a predictive role for rhythm in speech perception. ***The Quarterly Journal of Experimental Psychology, 38a***, 177-192.

Pitt, M.A., & Samuel, A.G. (in press). The use of rhythm in attending to speech. ***Journal of Experimental Psychology: Human Performance and Perception.***

Rackerd, B., & Fowler, C.A. (1984). The isochrony of speech does not appear to be syntactically conditioned. Paper presented at the meeting of the Michigan Linguistics Society, East Lansing, MI

Segui, J., & Frauenfelder, U. (1986). The effect of lexical constraints on upon speech perception. In F. Klix & H. Hagendorf (Eds.), ***Human memory and cognitive capabilities***. Amsterdam: North-Holland.

Segui, J., Frauenfelder, U., & Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. ***British Journal of Psychology, 72***, 471-477.

Selkirk, E.O. (1980). The role of prosodic categories in English word stress. ***Linguistic Inquiry, 11***, 536-605.

Selkirk, E.O. (1984). ***Phonology and syntax: The relationship between sound and structure.*** Cambridge: MIT Press.

Shields, J.L., McHugh, A., & Martin, J.G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. ***Journal of Experimental Psychology, 102***, 250-255.

Small, L.H. & Bond, Z.S. (1982, November). ***Effects of mispronounciation on lexical access in continuous speech perception*** . Paper presented at the meeting of the American Speech-Language Hearing Association, Toronto.

Suci, G.I., Ammon, P., & Gamlin, P. (1967). The validity of the probe-latency technique for assessing structure in language. ***Language and Speech, 10***, 69-80.

Swinney, D. (1984). Theoretical and methodological issues in cognitive science: A psycholinguistic perspective. In W. Kintsch, J.R. Miller & P.G. Poson (Eds.), ***Methods and tactics in cognitive science.*** Hillsdale, NJ: Lawrence Erlbaum Associates.

Walker, E.C.T. (1976). Some grammatical relations among words. In R.J. Wales & E.C.T. Walker (Eds.), ***New approaches to language processes.*** Amsterdam: North-Holland.

## Author Notes

The research reported in this paper was supported by grant 80-0416 to Harvard University (Peter C. Gordon, Principal Investigator) from the Air Force Office of Scientific Research Life Sciences Directorate. We thank Chris Thurber for his extensive and competent assistance in testing subjects and making acoustic measurements, and Jay Rueckl for comments on an earlier version of this paper

Direct correspondence to Peter C. Gordon, Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138.

Table 1

Examples of Sentence and Target Types Used in Experiments 1 & 2

AMBIGUOUS CONTEXT

1st Syll Stressed:

"Class ***CONflicts*** give rise to revolution."

2nd Syll Stressed:

"Class ***conFLICTS*** with my two noon appointments."

UNAMBIGUOUS CONTEXT

1st Syll Stressed:

"The ***CONflicts*** could not be resolved by conventional means."

2nd Syll Stressed:

"The discussion group often ***conFLICTS*** with other class meetings."

## Figure Captions

**Figure 1.** Mean response times for the syllable-monitoring task used in Experiment 1. The syllables next to the data points are example target syllables. First and second syllable refer to the position of the target syllable within the target-bearing word.

**Figure 2.** Mean response times and accuracies for the memory probe task used in Experiment 2. The numbers in parentheses indicate the mean accuracy for the associated response time.

## APPENDIX

### Stimuli Used in Experiments 1 and 2

Unambiguous Verb Condition

The FDA tries to **sub/JECT** new drugs to extensive animal testing.

The war is bound to **com/POUND** the junta's problems.

They should **per/MIT** the players to warm up.

The lawyer forgot to **ad/DRESS** the letter to her client.

Sears usually **re/FUNDS** shoppers for returned merchandise.

The outcome of the election may **up/SET** many people.

The inspectors often **re/JECT** over half of what they see.

The discussion group often **con/FLICTS** with other class meetings.

The committee will eventually **re/JECT** their decision.

Heroin rarely **ad/DICTS** first time users.

The candidate tried to **re/LAY** her fears of the deficit to voters.

Their army now **con/SCRIPTS** every healthy male between 18 and 20.

This couch can **con/VERT** into a queen-sized bed.

The Beastie Boys plan to **re/CORD** an album of punk-polka fusion music.

I have to **ob/JECT** to the way women are portrayed in the movies.

She usually **com/PACTS** her statements into a single paragraph.

It would be interesting to **con/TRAST** your style with Kunderas'.

The mayor secretly plans to **in/CREASE** taxes to balance the budget.

The patient tried to **con/TRACT** a lawyer to sue her doctor.

Ford announced plans to **re/CALL** over seven thousand cars.

The owner **ex/PLOITS** workers who don't have green cards.

He used a pair of needle-nosed pliers to **ex/TRACT** the gear.

Charlie **re/FILLS** these tanks every two weeks.

The farmers planned to **con/STRUCT** the barn in just three days.

The reader has to **ab/STRACT** the author's intent from her imagery.

Unambiguous Noun Condition

The student hates this **SUB/ject** but loves her other classes.

We created a new **COM/pound** for the project.

The hunters lost their **PER/mit** in the woods.

Her new **AD/dress** is in South Yarmouth.

I am hoping my **RE/funds** will arrive this week.

The Celtics win was a major **UP/set** for the Lakers.

We fixed up the **RE/ject** so it could be used as a spare.

The **CON/flicts** could not be resolved by conventional means.

The notorious J-walker did not show any **RE/gret** for his actions.

There are literally thousands of **AD/dicts** in this town.

The satellite picked up a **RE/lay** from the Cuban state news agency.

The young **CON/scripts** were sent to Levenworth for a minor offence.

My roommate has been a **CON/vert** to Cajun food since she tried my gumbo.

The Beastie Boys' new **RE/cord** is not as good as their last one.

The mysterious **OB/ject** showed up in a gully behind the building.

She must have left her **COM/pacts** in the dressing room after rehearsal.

There is quite a **CON/trast** between their personal styles.

The senators proposed a modest tax **IN/crease** to cover the cost of the plan.

Larry Bird signed a **CON/tract** to play with the Celtics for five years.

This is the largest **RE/call** in Ford history.

The explorer's **EX/ploits** are detailed in her new book.

I used vanilla **EX/tract** because vanilla beans are too expensive.

The waitress offered the customers **RE/fills** for their coffee.

The theoretical **CON/struct** Strauss presents is dubious at best.

I think I like the **AB/stract** more than I like the photorealist treatment.

Ambiguous Verb Condition

The psychology labs **sub/JECT** patients to extensive testing.

The Marcoses **com/POUND** Aquino's problems.

I hope the teams **per/MIT** us to enter.

The democrats **ad/DRESS** social issues briefly.

Caldors **re/FUNDS** shoppers five dollars for each bike they buy.

The candidates' debate **up/SET** most of the people who saw it.

The old **re/JECT** these silly ideas.

Class **con/FLICTS** with my two noon appointments.

Sometimes the old **re/GRET** missed opportunities.

Cocaine **ad/DICTS** thousands.

The White House communications **re/LAY** a sense of pessimism.

Their army **con/SCRIPTS** twelve thousand men a year.

The religious **con/VERT** uncertainty into faith.

The runners **re/CORD** their times.

My speeches **ob/JECT** to the use of nuclear power.

The new Izusu **com/PACTS** the snow in the drive.

My uncle Jim's new t.v.'s **con/TRAST** control knob is missing.

Tax **in/CREASES** the cost of gas.

The patients **con/TRACT** a respiratory disease if left untreated.

The automakers **re/CALL** defective cars from time to time.

The mercenary **ex/PLOITS** his connections with the deposed junta.

The Wampanoag indians **ex/TRACT** purple dye from sea shells.

Dunkin Donuts **re/FILLS** your coffee cup for free.

My philosophy professors **con/STRUCT** treehouses in their spare time.

The physics papers **ab/STRACT** new predictions from Einstein's laws.

Ambiguous Noun Condition

The psychology lab's **SUB/ject** is a 34 year old schizophrenic male.

The Marcos's **COM/pound** surrounds the beach front.

I hope the teams' **PER/mit** is valid.

The democrats' **AD/dress** is short on specifics.

Caldors' **RE/funds** come to five dollars on each new bike.

The candidate's debate **UP/set** put her ahead in the polls that week.

The old **RE/ject** was placed in the bin.

Class **CON/flicts** give rise to revolution.

Sometimes the old **RE/gret** is left unmentioned.

Cocaine **AD/dicts** are sick.

The White House communications **RE/lay** system was struck by lightening.

The army **CON/scripts** were disadvantaged youths.

The religious **CON/vert** is eager to proselytize.

The runner's **RE/cord** still stands.

My speech's **OB/ject** is to convince people to quit smoking.

The new Izusu **COM/pacts** have power steering.

My uncle Jim's new t.v.s' **CON/trast** control knob seems to be missing.

Tax **IN/creases** are unpopular.

The patient's **CON/tract** does not cover long-term hospitalization.

The automakers' **RE/call** may eventually cost Detroit two thousand jobs.

The mercenary **EX/ploits** of Col. North did not impress the judge.

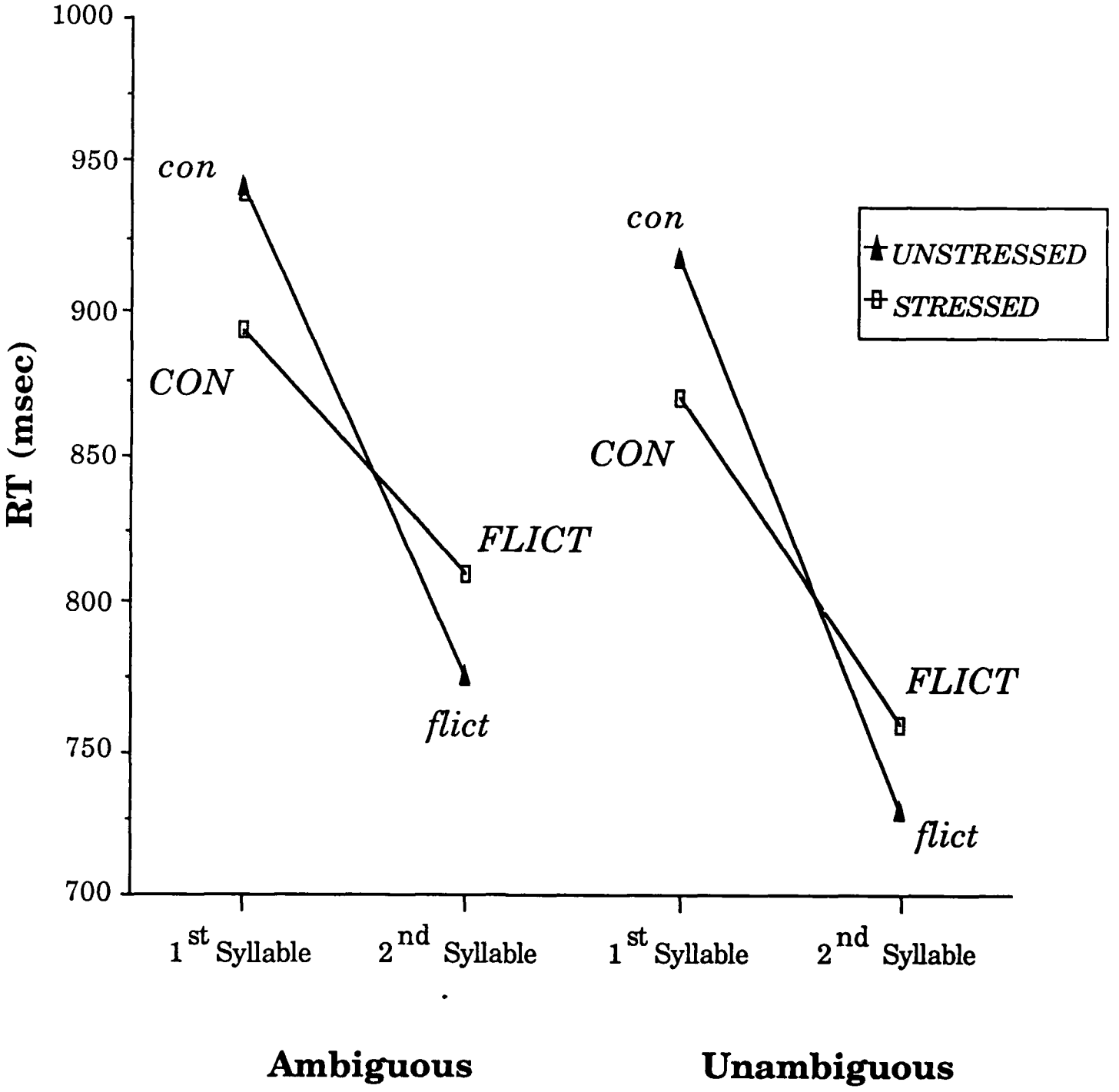
The Wampanoag Indian's **EX/tract** is kept in a ceremonial conch shell.

Dunkin Donut's **RE/fills** of coffee are free.

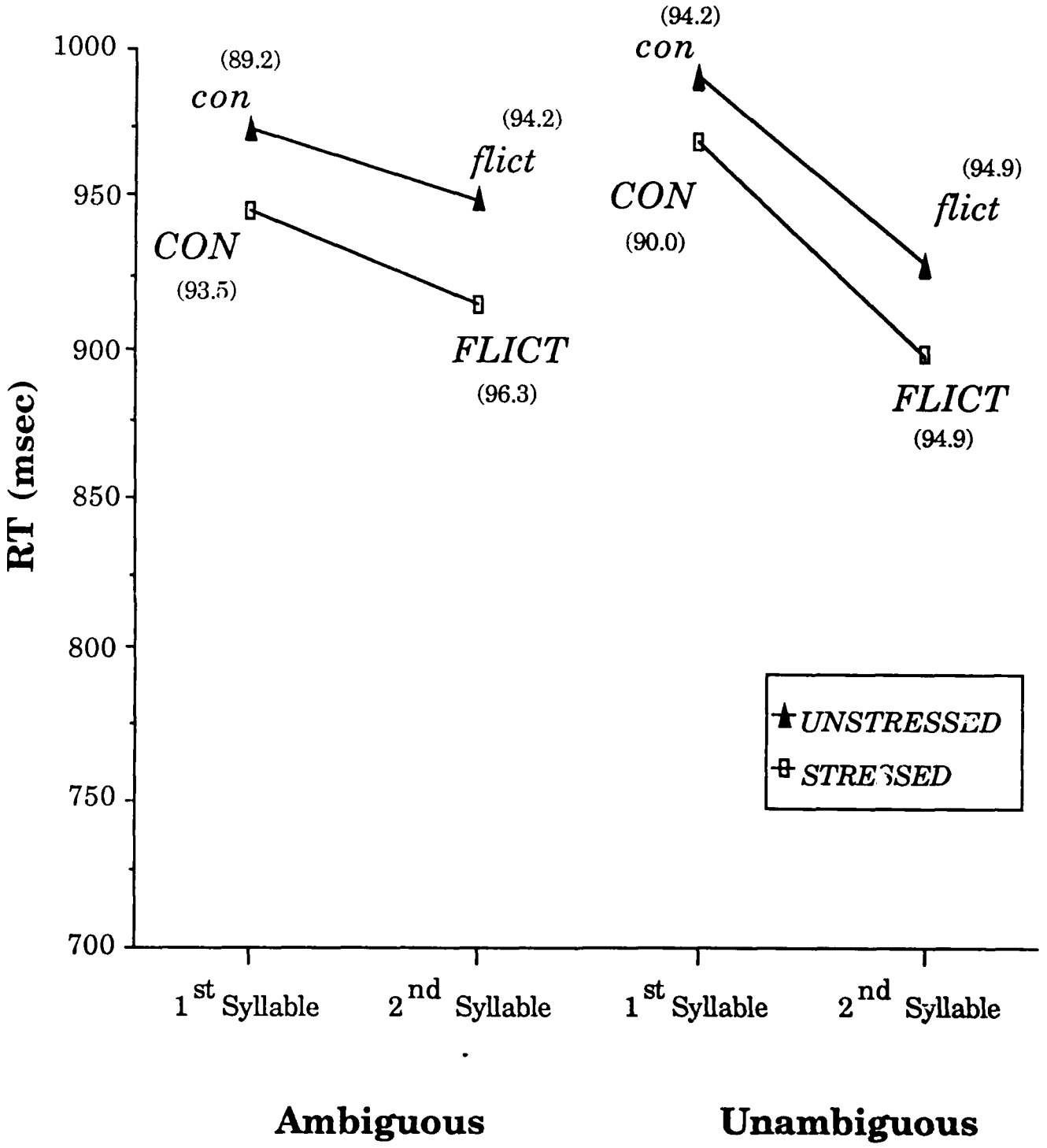
My philosophy professor's **CON/struct** is inherently flawed.

The physics paper's **AB/stract** was hard to follow.

# Figure 1


1000
950
900
850
800
750
700
RT (msec)
con
CON
FLICT
flict
con
CON
FLICT
flict
UNSTRESSED
STRESSED
1st Syllable
2nd Syllable
1st Syllable
2nd Syllable
Ambiguous
Unambiguous

# Figure 2


1000
950
900
850
800
750
700
RT (msec)
(89.2)
con
(94.2)
flict
CON
(93.5)
FLICT
(96.3)
(94.2)
con
(94.9)
flict
CON
(90.0)
FLICT
(94.9)
UNSTRESSED
STRESSED
1st Syllable
2nd Syllable
1st Syllable
2nd Syllable
Ambiguous
Unambiguous